

NORTH ATLANTIC TREATY ORGANIZATION



RESEARCH AND TECHNOLOGY ORGANIZATION

BP 25, 7 RUE ANCELLE, F-92201 NEUILLY-SUR-SEINE CEDEX, FRANCE

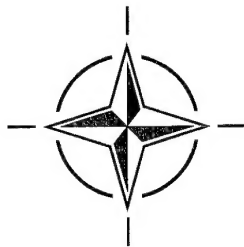
RTO LECTURE SERIES 216

Application of Mathematical Signal Processing Techniques to Mission Systems

(l'Application des techniques mathématiques du traitement du
signal aux systèmes de conduite des missions)

*The material in this publication was assembled to support a Lecture Series under the
sponsorship of the Systems Concepts and Integration Panel (SCI) presented on 1-2 November
1999 in Köln, Germany, on 4-5 November 1999 in Paris, France, and on 9-10 November 1999
in Monterey, USA.*

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited



19991130 072

Published November 1999

Distribution and Availability on Back Cover

The Research and Technology Organization (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote cooperative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective coordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also coordinates RTO's cooperation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of initial cooperation.

The total spectrum of R&T activities is covered by 7 Panels, dealing with:

- SAS Studies, Analysis and Simulation
- SCI Systems Concepts and Integration
- SET Sensors and Electronics Technology
- IST Information Systems Technology
- AVT Applied Vehicle Technology
- HFM Human Factors and Medicine
- MSG Modelling and Simulation

These Panels are made up of national representatives as well as generally recognised 'world class' scientists. The Panels also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier cooperation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

The content of this publication has been reproduced directly from material supplied by RTO or the authors.



Printed on recycled paper

Published November 1999

Copyright © RTO/NATO 1999
All Rights Reserved

ISBN 92-837-1021-5



*Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7*

Application of Mathematical Signal Processing Techniques to Mission Systems

(RTO EN-7)

Executive Summary

Signal processing techniques must develop substantially, on the one hand in order to respond in a more relevant way to more demanding operational requirements, and on the other to obtain maximum benefit from improvements in the technologies on which they are based, whether it be for the sensors which supply them, or the data processing techniques which enable their implementation.

With regard to sensors in particular, the trend is to use the signal for imaging, at increasingly fine resolution, with generally much larger fields. Moreover, processing commonly concerns sequences of images, with close integration of spatial and temporal dimensions. Present day systems in fact tend to multiply the number of sensors and frequency bands operated in close synergy, leading to multi-resolution and non-uniform data (reference systems, reliability,...). The data available are thus increasing in volume, in density and in irregularity, and as a result are becoming more difficult to use.

Operational situations require the generation of increasingly accurate, undeformable and summarised information, to be generated under more and more difficult conditions with shorter and shorter reaction times. The data and the interconnections which result from it, must therefore be treated with care, while at the same time attempting to ensure the highest possible level of automaticity.

There are a number of emerging techniques which could meet these requirements, mostly originating in mathematical theories as diverse as wavelets, variational methods or the theory of evidence. These techniques cover the whole processing chain fairly evenly, and in particular signal compression and transmission, data extraction and interpretation, and decision-making aids.

JUSTIFICATION: The complementarity of the different emerging techniques, presented in the most varied mathematical frameworks, so as to respond to what is a critical development in sensor system integration requirements, should produce a series of tools capable of meeting the needs expressed at all levels of the processing chain.

SUBJECTS EXAMINED: This Lecture Series presents a whole range of perspectives for different levels of processing, based on some of the most promising techniques. Particular attention will be paid to the following subjects:

- Wavelet analysis: summary of the possibilities; application to detection in natural background radiation and extraction of primitive invariants.
- The concept of Multirate Filter Banks in conjunction with the various transforms which this technique enables; applications to compressed video image and sequence transmission, to noise rejection, to jamming and to encoding.
- Variational methods based on partial derivative equations for image processing and multi-scale video sequences; presentation of different image segmentation approaches.
- Multi-sensor processing based on the theory of evidence: processing of the functions of detection, classification, matching of ambiguous observations, or tracking, with the aim of solving problems such as data modelling, decision making, the management of non-uniform reference systems, or the integration of contextual knowledge.

The material in this publication was assembled to support a Lecture Series under the sponsorship of the Systems Concepts and Integration Panel (SCI) and the Consultant and Exchange Programme of RTA presented on 1-2 November 1999 at DLR Koln, Germany, on 4-5 November 1999 at ONERA, Paris, France, and 9-10 November 1999 at the Naval Post Graduate School, Monterey, United States.

L'application des techniques mathématiques du traitement du signal aux systèmes de conduite des missions

(RTO EN-7)

Synthèse

Les techniques de traitement du signal doivent évoluer de façon substantielle, d'une part pour répondre d'une façon pertinente à des besoins opérationnels de plus en plus exigeants, et d'autre part pour tirer tout le bénéfice de l'amélioration des technologies sur lesquelles elles reposent, qu'il s'agisse des senseurs qui les alimentent ou des moyens informatiques qui permettent leur mise en œuvre.

Au niveau des senseurs en particulier, le signal évolue de plus en plus vers l'imagerie dont la résolution est de plus en plus fine pour des champs généralement plus importants. Il faut traiter le plus souvent, des séquences d'images et ceci en intégrant étroitement leurs dimensions temporelle et spatiale. Les systèmes actuels multiplient de plus le nombre de senseurs et de bandes de fréquence qu'il convient d'exploiter en étroite synergie, conduisant notamment à des problèmes de multi-résolutions et d'hétérogénéité des données (référentiels, fiabilité,...). Les données disponibles croissent donc en volume, en richesse, en hétérogénéité, et en difficulté d'exploitation.

Les besoins opérationnels requièrent par ailleurs l'élaboration d'informations de plus en plus précises, robustes, synthétiques, ceci dans des conditions adverses souvent plus difficiles et avec des délais de réaction de plus en plus courts. Il convient donc d'exploiter de façon d'autant plus rigoureuse les données et leurs synergies, tout en cherchant un niveau d'automatisation le plus élevé possible.

Pour faire face à ces besoins, un certain nombre de techniques émergentes et porteuses ont pu être dégagées à partir de théories mathématiques aussi variées que les ondelettes, les méthodes variationnelles ou la théorie de l'évidence. Ces techniques couvrent de façon assez homogène l'ensemble de la chaîne de traitement, notamment la compression et la transmission des signaux, l'extraction d'information, l'interprétation, et l'aide à la décision.

JUSTIFICATION : Les complémentarités de différentes techniques émergentes et porteuses, élaborées dans des cadres mathématiques les plus variés pour répondre à une évolution critique des besoins en matière d'intégration de systèmes de senseurs, permettent d'envisager un ensemble d'outils propres à satisfaire tous les maillons de la chaîne de traitement.

SUJETS A TRAITER : Le cycle de conférences proposé vise à présenter un éventail des perspectives offertes aux différents niveaux du processus de traitement, en s'appuyant sur quelques techniques parmi les plus prometteuses. Les sujets suivants seront notamment abordés :

- Analyse par ondelettes : synthèse des possibilités offertes ; application à la détection dans des fonds naturels structurés et à l'extraction de primitives invariantes ;
- Concept de "Multirate Filter Banks" en liaison avec les différentes transformées qu'il permet de mettre en œuvre ; applications dans le domaine des transmissions à la compression d'images et de séquences vidéo, à la réjection de bruit, au brouillage, et au codage ;
- Méthodes variationnelles basées sur les équations aux dérivées partielles pour le traitement d'images et de séquences vidéo multi-échelles ; présentation de différentes approches en segmentation d'images ;
- Traitements multi-senseurs basés sur la théorie de l'évidence : traitement des fonctions de détection, classification, mise en correspondance d'observations ambiguës, ou pistage, visant à résoudre des problèmes tels que la modélisation des données, la prise de décision, la gestion de référentiels hétérogènes, ou l'intégration de connaissances contextuelles.

Cette publication a été rédigée pour servir de support de cours pour le Cycle de conférences 216, organisé par la Commission RTO sur les (SCI) du 1-2 novembre 1999, DLR, (Allemagne) et du 4 au 5 novembre 1999 à l'ONERA, (France), et du 9 au 10 novembre 1999 à Naval Post Graduate School, Monterey (Etats-Unis).

Contents

	Page
Executive Summary	iii
Synthèse	iv
List of Authors/Speakers	vi
	Reference
Introduction to Wavelet Analysis by G.H. Watson	1
The Detection of Unusual Events in Cluttered Natural Backgrounds by G.H. Watson	2
Invariant Feature Extraction in Wavelet Spaces by G.H. Watson	3
Multirate Filter Banks and their Use in Communications Systems by C.D. Creusere	4
Multisensor Signal Processing in the Framework of the Theory of Evidence by A. Appriou	5
Partial Differential Equations for Multiscale Image and Video Processing by G. Hewer and C. Kenney	6

List of Authors/Speakers

Lecture Series Director:

Mr Robert W. CAMPBELL
Deputy, Weapons and Targets Dept.
Code 47A
Navairwarcentwpndiv
1 Administration Circle
China Lake CA 93556-6100
USA

AUTHORS/LECTURERS

Dr Charles CREUSERE
Code 4T4400D
Naval Air Warfare Center Weapons Division
1 Administration Circle
China Lake CA 93555-6100
UNITED STATES

Dr Gary HEWER
Code 471600D
Naval Air Warfare Center
1 Administration Circle
China Lake CA 93555-6100
UNITED STATES

Mr Alain APPRIOU
ONERA-DTIM
29 Av de la Division Leclerc
92322 Chatillon Cedex
FRANCE

Mr Graham WATSON
Room 1052, A2 Building
DERA Farnborough
Ively Road
GU14 0LX
UNITED KINGDOM

CO-AUTHORS

Mr Charles KENNEY
ECE Department
University of California
Santa Barbara, CA 93106
UNITED STATES

Introduction to Wavelet Analysis

G.H. Watson

Room 1052, A2 Building, DERA Farnborough, Ively Road, Farnborough, Hants, GU14 0LX, UK

1. Introduction

This paper introduces the concepts of wavelet analysis and gives an overview of the numerous wavelet analysis techniques in existence. The principal aim of this paper is to promote an awareness of wavelet analysis, not to provide technical details, as the latter are available in many textbooks, for example [1,2]. Most of the underlying principles are applicable to 1-dimensional signal analysis, and there are straightforward methods to adapt 1D wavelet analysis to higher-dimensional data, also covered in this paper. Hence, much of this paper is concerned with 1-dimensional signal analysis, even though higher-dimensional data is of equal importance. Major topics covered in this paper are the continuous wavelet transform and its inverse, the discrete wavelet transform and its relation to multiresolution filter banks, orthonormal and biorthogonal wavelets, image wavelet analysis and wavelet packets.

We begin in this section with an overview of what wavelet analysis is, why it is useful, and present some common applications. Throughout this paper, key words and phrases are highlighted in bold text.

Wavelet analysis is the extraction of signal or image information at different **positions** and **scales**. The idea is to treat all positions and scales on an equal footing, so that an object will be analysed in the same way, regardless of whether it is translated or dilated. This approach is useful because **translation** and **dilation** are natural symmetries that occur very often in nature, and in signal and image processing. If we are looking for an object, we generally don't know where it will be, and in many surveillance applications it is equally likely to be anywhere in the signal or image. The statistics of the signal or image are thus translation-invariant, otherwise known as being stationary. Similarly, if we're analysing signals over time, we don't know when an event will occur, for example a transient sound in an acoustic signal.

Scale invariance is also important in signal and image processing, but the reasons are sometimes less obvious. Sometimes scale-invariant processing is required because the objects being analysed could be at any range, and therefore of unknown apparent size, or the camera may have a zoom facility which also dilates the image. Similarly, sounds such as musical notes may have variable duration, but in other respects are similar. What is more subtle and interesting is the invariance of

many natural processes and scenes to dilation. Scenes such as sky, clouds, mountains and forests are of interest as backgrounds in surveillance and detection. It should be obvious that such backgrounds are statistically independent of translation, as there is no concept of "absolute" position. This is similar to the underlying principle of relativity, although the latter concerns the laws of physics and also invariance to constant velocity changes.

What is less obvious is that many natural scenes are scale-invariant; when we observe such scenes as images, the range or magnification are difficult to discern, unless there are reference objects of known size. Even many artefacts, such as roads and buildings, are difficult to scale. Self-similar objects are known as fractals, and the study of fractal geometry has been an important topic of research in recent decades [3], in which scale-invariance is known as **self-similarity**. There are many physical processes which are self-similar, for example turbulence in fluids, and wavelet analysis has been an important tool in the analysis of such processes.

There are natural symmetries other than translation and dilation, which will be mentioned in Sections 7 and 8. Downward-looking imagery is often statistically rotation-invariant, there being no bias in orientation. Frequency shifts are a natural symmetry for some types of noise, for example Gaussian white noise.

Another important requirement of wavelet analysis is **resolution** in position and scale, so that objects at different positions and scales can be analysed independently, with minimal interference. To achieve this, an appropriate **basis** of functions is required for the analysis. The most primitive basis comprises the delta functions which return the sample or grey-scale values at each point or pixel in the signal or image. Delta functions are best at resolving position but cannot resolve scale or frequency. Conversely, a Fourier basis, comprising sinusoids or complex exponentials, is best at resolving frequency, but cannot resolve position. Neither of these bases is scale-invariant, which is where wavelet bases come in, discussed in Section 2.

We conclude this section with some applications of wavelet analysis, to demonstrate the practical importance of translation- and scale-invariant processing.

1.1 Data Compression

Data compression is perhaps the most widely used application of wavelet analysis. Most real-life images have strong phase correlation, like edges, and are intermittent, with some parts being smooth and other parts rough, or with sharp edges. With the delta function basis there is considerable redundancy in the smooth parts, as the function values (sample values) are similar, so smooth regions require a basis of smooth functions to be encoded efficiently. There is a lot of low-frequency energy in smooth signals, which suggests that a Fourier Transform might be more efficient, but sharp edges are a problem, because they have energy over a wide range of frequency. Thus we would need to partition the image into regions each with separate frequency decomposition, which leads to windowed Fourier (or cosine) transforms, for example the discrete cosine transform (DCT) used in JPEG image compression [4]. A similar technique is used in encoding audio signals in the form of the Gabor transform or spectrogram [5]. Thus edges can be decomposed separately, leaving smooth regions to be encoded more efficiently.

The windowed Fourier technique is quite effective, but this type of coding is still limited because a fixed window size is used. If a large window is used, edges and high frequency energy are coded badly, because there is significant leakage into smooth regions, as windows of fixed size and regular spacing do not usually fit edges well. If small windows are used, low frequency smooth regions are coded badly, as there are too many windows replicating information. What we need is a variable-scale window, which is where the wavelet transform comes in. The above coding problems are caused by a lack of scale invariance, as a fixed window does not treat different scales alike.

If we use the wavelet transform, the signal or image is decomposed into a **pyramid**, each layer having information at a different scale and level of detail. Each layer comprises a regular grid, where at each point there is a wavelet coefficient encoding the information within the image at that particular position and scale. The grid spacing is proportional to scale, so a small number of coefficients is required at large scale and low resolution. Thus smooth parts of the image are encoded efficiently. At small scales a large number of coefficients is required, but in smooth areas these will be low in magnitude, and can be ignored with minimal loss of information. Thus we are getting what we want: smooth regions are encoded with a small number of coefficients, and other regions, such as edges, are encoded with a larger number of coefficients. Fig. A gives an example of image compression using symmetrical Daubechies wavelets.

Fig. A. Example of wavelet image compression on 'Lena'



(a) Original image



(b) Image at 27:1 compression

1.2 De-Noising

If a signal or image is corrupted with noise, we wish to recover as much of the original information as possible. We cannot do a perfect job, because some parts of the signal will be indistinguishable from noise; they could have arisen with some probability from the random process generating the noise. The usual method is to decompose the signal into a set of functions using a prescribed basis (in our case using a wavelet basis), distinguish the components that come from noise from those that don't (to some level of confidence), remove the former, and reconstruct the signal or image from the latter. The role of the basis is to do the best possible job of separating the original signal and noise. The best choice of basis depends both on what we expect to find in the uncorrupted signal, and on the statistical properties of the noise.

When the expected signal is self-similar both in position and scale, the wavelet transform is the obvious method of decomposition. If we have Gaussian white noise then it turns out that the resulting wavelet coefficients all have the same Gaussian distribution, so the natural way of de-noising is to set a threshold on the amplitudes of the wavelet coefficients, and to reject (set to zero) all those below this threshold. If, say, the probability of the wavelet coefficients from the noise exceeding this threshold is only 1/1000, then anything remaining is more than 99.9% likely to come from the signal. There is a trade-off between missing too much of the signal and leaving too much of the noise, and the required balance affects the value of the threshold. Fig. B gives an example of signal de-noising using wavelet analysis.

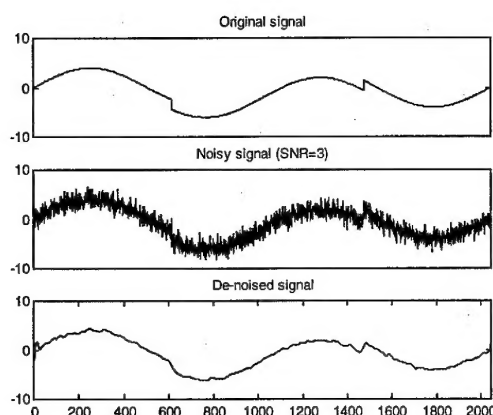


Fig. B. Example of de-noising of a 1-dimensional signal using wavelet analysis

1.3 Detection

Finally we briefly consider target and anomaly detection, which is covered in more detail in [6]. Detection is very similar to de-noising, except now we may not need to reconstruct the uncorrupted data. It is therefore often sufficient to record the position, scale and amplitude of the wavelet components, and so an inverse to the wavelet transform is not necessary. This gives us more flexibility in the choice of the wavelet (or other) basis, not just in the shape of the functions, but also in their spacing in positions and scale. Typically we can afford to choose a higher threshold on the wavelet coefficients, and to use a denser pyramid, thus over-sampling the wavelet transform. This extra processing, discussed in the last two sections, allows better target discrimination, but can also introduce redundancy in the representation of the target.

2. Fundamentals: the Continuous Wavelet Transform

This section introduces wavelet analysis for continuous functions, where the concepts of translation and dilation

are clearest. Sections 3–5 cover the analysis of discretely sampled signals.

2.1 Convolution

The wavelet transform is essentially a multi-scale convolution of a signal with a filter, called the **analysing wavelet** or **mother wavelet**. First we briefly review single-scale convolution. Convolution of a signal f with a filter g is defined as follows:

$$h = f * g \quad ; \quad h(x) = \int_{\mathbb{R}^n} f(u)g(x-u)du \quad (1)$$

where integration is over the space on which the functions f and g are defined. If f is a function of one variable, e.g. an acoustic signal is a function of time, then so must be the filter g , and the integral is one-dimensional, i.e. on the real line. The convolution output h is also a function of one variable: x is a scalar quantity. For image processing f is a function of two variables, so x and u are vectors, each with two scalar components. The integral is two-dimensional, i.e. over the image plane. Convolution can also be done in higher dimensions, for example when analysing time-sequenced imagery or medical tomography.

In both cases the underlying principle is the same: we take a filter function g , reverse it in space or time, and slide it over the signal f over all positions x , which is done by translating g , and is why the argument of g under the integral is $x-u$, not u . For images the translation is a vector, allowing the filter to be positioned anywhere within the image. The value of the convolution output h at x tells us how the signal or image interacts with the filter at that particular position x .

Convolution is translation invariant; if the signal is shifted, then the convolution output is shifted by the same amount. Thus convolution is a natural precursor to wavelet analysis, suitable for analysing signals with translation invariance, where the information sought is equally likely to occur at any position (or time). However, convolution does *not* treat scales on an equal footing; if the signal is dilated, the convolution output is not dilated or simply related in any other way. Table 1 shows some simple examples of convolution filters.

The top hat is a local average, so it integrates the signal over an interval of unit length, and the output of h depends on the starting point. This will be good at identifying regions in the signal with high (or low) local average, for example a pulse, but will also respond well to signals with a high *global* average, for example a constant non-zero function. Thus it will be good at discriminating pulses from the background so long as the local mean of the latter is always small, which requires the background to be uncorrelated over lengths comparable to the scale of the filter, for example zero-mean white noise. In this case it will be better at picking

up pulses of approximately unit length than of much smaller or longer lengths because the signal to noise ratio is higher. This is the principle behind matched filters. The important point is that the effectiveness of the filter depends on the scale of the object it is trying to detect.

The Gaussian pulse is similar to the top hat, but being smoother it is less sensitive to high frequencies, and thus better at picking up smoother objects. The edge detector is rather different, as it only responds to changes or gradients within the signal, because anything constant is cancelled out by the up and down pulses: the filter has zero-mean. Thus this is a good edge detector, especially if the background is highly correlated, e.g. Brownian noise, because, looking for differences only, it ignores highly correlated regions. The same thing goes for the Mexican hat or Difference of Gaussian (DoG) filter, except it is symmetrical, and responds best to 2-sided edges (filaments in images). Again, there is a scale-dependence; in Brownian noise these filters respond best to smooth ramps whose width is approximately unity.

2.2 The Wavelet Transform

The wavelet transform removes this scale-dependence by repeating the convolution of Equation (1) at multiple scales, producing a function of position and scale:

$$w(x, s) = s^{-\frac{n}{2}} \int_{\mathbb{R}^n} f(u) g\left(\frac{x-u}{s}\right) du \quad (2)$$

so that the filter is dilated by a factor s as well as translated by an offset x . The power of scale in front of the integral is a normalisation factor similar to the

factors involving π used in the Fourier transform. One useful property of this normalisation factor, discussed in Section 8, is that the expected wavelet transform of white noise is independent of scale. Now all information is treated similarly, regardless of position and scale. Any translation and dilation of the signal or image will result in a similar translation and dilation of the wavelet transform. The filter g is known as the **analysing wavelet** or **mother wavelet**, and depending on its shape (e.g. Table 1), the wavelet transform will be good at detecting top hats, pulses and edges at all positions and scales.

2.3 Inverse Wavelet Transform and Admissibility

As you would expect for a useful transform, there is an inversion formula:





$$f(x) = C^{-1} \int_{\mathbb{R}^{n+1}} w(u, s) g\left(\frac{x-u}{s}\right) s^{\frac{3n}{2}} du ds \quad (3)$$

where C is a normalisation constant given by:

$$C = 2\pi \int_{\mathbb{R}} \frac{\hat{g}(\omega)}{\omega} d\omega, \quad (4)$$

where the hat denotes a Fourier transform and ω is Fourier frequency. This formula is analogous to the continuous Fourier transform inverse, in that both transforms look very similar to their inverses, and indeed, the wavelet inverse is easiest to derive in the Fourier domain, using the Fourier inversion theorem. The wavelet transform inverse is more powerful, because it works for a large family of mother wavelets, in fact any function g for which the normalisation

Table 1. Example convolution filter functions.

Name	Function	Approximate shape
Top hat	$g(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$	
Gaussian pulse	$g(x) = \exp(-x^2)$	
Simple edge detector	$g(x) = \frac{d}{dx}(\exp(-x^2))$	
Mexican hat (DOG)	$g(x) = \frac{d^2}{dx^2}(\exp(-x^2))$	

constant C is finite, whereas the continuous Fourier transform involves convolution with complex exponentials only. The finiteness of C imposes a significant constraint on g however, called the **admissibility condition**, in particular requiring g to have zero mean. Thus the inversion formula (3) does not work with the top hat and Gaussian pulse functions in Table 1. Many practitioners of wavelet analysis require the admissibility condition as part of the definition of a wavelet. However, the wavelet transform (2) still has meaning, and translation and dilation invariance, even without this condition; it is mainly when using the inversion formula (3) that the admissibility condition is required.

3. Discrete Wavelets and Filter Banks

3.1 The Effects of Sampling

The continuous wavelet transform is sound theoretically, but it not applicable to signal and image analysis with digital computers, which require discretely sampled data, discrete filters, and where integration is replaced with finite summation. The same argument applies to Fourier analysis, which is why in practice the discrete Fourier transform is used, often implemented as the fast Fourier transform (FFT). Similar implementations have been developed for wavelet analysis, and there is an elegant relationship between the continuous and discrete cases, described in Section 4.

We require discrete equivalents for the operations shown in Table 2. Dilation is the main cause of difficulty, and the reason for various complications in the theory of the discrete wavelet transform, because downsampling and upsampling are not invertible even though they appear superficially to be inverses of each other. It is true that upsampling followed by downsampling is the identity, leaving the signal unchanged, but if these operations are applied in the reverse order all the samples whose index k is not divisible by p are set to zero, and thus information is lost.

3.2 Filter Banks and Perfect Reconstruction

We need to avoid losing information, otherwise the discrete wavelet transform will not be invertible, and the signal or image would not be fully represented. For this reason it is necessary to apply more than one discrete filter to the data, in fact p filters, where p is the resampling factor. Thus discrete wavelet analysis involves the application of filter banks. Fig. 1 shows the process, involving a single dilation and its inverse, in diagrammatic form.

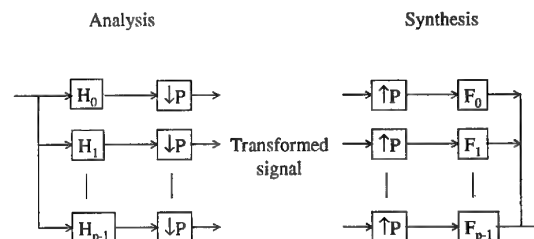


Fig. 1. Signal analysis and synthesis

If the reconstructed signal coming from the synthesis channel is identical, barring a delay, to the input to the analysis channel, the filter bank is called a **perfect reconstruction (PR)** filter bank. H_k are called **analysis filters** and F_k are called **synthesis filters**. Both are discrete, linear and translation invariant (to a resolution of one sample), and in general are implemented recursively:

$$h(n) = \sum_k a(k)h(n-k) + \sum_l b(l)f(n-l) \quad (5)$$

where the coefficients $a(k)$ and $b(l)$ are finite and their number defines the order of the filter. All such filters can be implemented by discrete convolution, where there are no recursive coefficients $a(k)$, but there may be infinitely many $b(l)$. The latter may be obtained by applying the filter to a delta function, or impulse, and hence are

Table 2. Continuous and discrete operation analogues

Continuous	Discrete	Discrete Formula
Integration	Summation	$\sum_k a(k)$
Translation	Shift to left or right by integer p	$f(k) \rightarrow f(k-p)$
Dilation	Downsampling or upsampling by integer factor p	$f(k) \rightarrow \begin{cases} f(k) & \rightarrow f(pk) \\ f(k/p) & k/p \text{ integer} \\ 0 & \text{otherwise} \end{cases}$

denoted the **impulse response**. Where the response is infinite, the filter is known as infinite impulse response (IIR), otherwise finite impulse response (FIR).

The other components of the filter bank are to do with resampling, where $\downarrow P$ denotes downsampling, and $\uparrow P$ denotes upsampling, both by a factor of p . Thus this filter bank has the discrete analogues of both convolution and dilation, and thus contains all the ingredients required for wavelet analysis. Most (though not all) resampling in wavelet analysis is by a factor of $p=2$, because like the fast Fourier transform the process is most efficient this way, so from this point we will assume $p=2$.

When designing filters, including PR filters, it is convenient to use the Z-transform, where the coefficients a and b are each assembled into a polynomial, and the transform is the ratio of these polynomials, where the independent variable is conventionally written as z . In this representation, the following conditions are necessary and sufficient for perfect reconstruction:

$$\begin{aligned} F_0(z)H_0(z) + F_1(z)H_1(z) &= 2z^{-1} \\ F_0(z)H_0(-z) + F_1(z)H_1(-z) &= 0 \end{aligned} \quad (6)$$

In this notation, multiplication is equivalent to composition of filters, changing the sign of z is equivalent to reversing the filter, and z^l is equivalent to a delay of l samples. The first equation ensures zero distortion and the second prevents aliasing.

3.3 Multiple Resolution: Discrete Wavelet Transform

We now have two channels in our analysis and synthesis (reconstruction) filter banks. The idea is to decompose the signal into components at multiple resolutions, with octave dilation factors (in general powers of p). We lower the resolution, i.e. shrink the signal or image, by downsampling, and reconstruct by upsampling. To work with multiple resolutions, we apply the filter bank in Fig. 1 recursively:

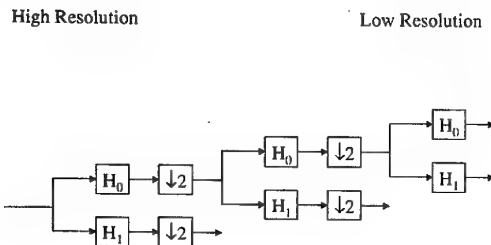


Fig. 2. Multiresolution Filter Bank

As the object is to shrink the image, it is conventional to apply a smoothing (lowpass) filter, to avoid aliasing. By convention, therefore, H_0 is a lowpass filter and H_1 is a highpass filter, so that all information about the signal or image is retained. Eventually we are left with just a small number of coefficients at the lowest resolution (in the limit just one), and a pyramid of highpass output values at multiple resolutions, each resolution (for 1D signals) having half the number of coefficients of the previous resolution. The case is slightly different for higher dimensions, for example images, covered in Section 6. This representation is the **discrete wavelet transform**, sometimes called a **pyramid**, because the number of coefficients decreases at each new level. The reconstruction of the signal from the pyramid is also done recursively in reverse order.

The highpass outputs are often referred to as **detail** coefficients, because they effectively siphon information at a particular resolution, the lowpass coefficients going to the next level. At each resolution, the combination of a number of iterations of H_0 , followed by H_1 can be regarded as a bandpass filter, and is the analogue of convolution of the signal with the analysing wavelet at the appropriate scale.

The method of convolution is rather different, however. In Equation (2) the convolution is done by dilating the filter g but keeping the signal f fixed. In the discrete implementation the filter is fixed and the signal is dilated by the inverse factor. The two operations are equivalent in the continuous case, as can be seen by substituting u/s for u in Equation (2), but in the discrete case they are not, because upsampling and downsampling are not inverses.

The main reason for shrinking the signal rather than expanding the filter is efficiency, as in the former case the computation decreases with resolution, whereas in the latter case the computational load increases, as the filter coefficients increase in number. However, there is a drawback. Downsampling means that the wavelet transform is evaluated on an increasingly sparse grid as the resolution decreases. This grid is prescribed, so the discrete wavelet transform is *not* translation invariant. If we shift the signal by k , the wavelet transform is translation-invariant only at resolutions which divide into k perfectly. Eventually the dilation factor will exceed k , and so translation invariance will break down. The wavelet transform of the translated signal effectively "falls between" points in the grid above a certain scale. The discrete wavelet transform is only scale-invariant for octave (powers of p) scale changes, for similar reasons. This failing has implications for detection and classification of objects, as they do not necessarily result in the same signature in the discrete wavelet transform when translated or dilated. The continuous wavelet transform does not have these disadvantages.

4. The Dilation and Wavelet Equations

4.1 Wavelets and Filter Banks

The multiresolution filter bank of Fig. 2 is essentially how the discrete wavelet transform is implemented, but the relationship with the continuous wavelet transform is rather loose, based on the analogue between dilation and resampling. Under certain conditions however, described in this section, there is a much stronger link between wavelets and filter banks, discovered by Stephane Mallat, summarised next. It is based on a construction which allows *continuous* wavelet transform coefficients to be computed using multiresolution filter banks. This is done using two equations: the **dilation equation**:

$$\phi(x) = \sum_k 2h_0(k) \phi(2x-k) \quad (7)$$

and the wavelet equation:

$$w(x) = \sum_k 2h_1(k) \phi(2x-k) \quad (8)$$

where h_0 and h_1 are the impulse response coefficients of the filters H_0 and H_1 respectively (the equivalent of the coefficients b in Equation (5) if the coefficients a are all zero). ϕ is called the **scaling function** and w is the **analysing wavelet** which has the same role as g in Equation (2). When convolved with the signal, ϕ acts as a lowpass filter (in much the same way as the discrete filter H_0) and w is a bandpass filter (analogous to H_1). Equations (7) and (8) allow the wavelet transform at one scale to be calculated from the same transform at half this scale, without direct convolution, using the discrete filters H_0 and H_1 . To see how this works, substitute ϕ and w for g in equation (2) to produce two functions $w_0(x,s)$ and $w_1(x,s)$, and then sample these functions on a discrete grid with octave scales and position spacing proportional to scale (pyramid sampling):

$$\begin{aligned} a(p,q) &= w_0(2^q, p) \\ b(p,q) &= w_1(2^q, p) \end{aligned} ; \quad p, q \text{ integers} \quad (9)$$

We then have the following recurrence relations for a and b :

$$\begin{aligned} a(p,q) &= \sum_k \sqrt{2} h_0(k-2p) a(k,q+1) \\ b(p,q) &= \sum_l \sqrt{2} h_1(k-2p) b(l,q+1) \end{aligned} \quad (10)$$

which is identical to using the coefficients $a(p,q)$ as input data into the filter bank of Fig. 1 to produce $a(p,q+1)$ and $b(p,q+1)$ as outputs. The multiresolution filter bank of Fig. 2 will therefore produce the values of the wavelet transform function w_1 on the pyramid grid, but much more efficiently than by direct convolution.

The filter bank in Fig. 2 works because the scaling function and analysing wavelet are carefully designed so that these functions can be dilated by translation, scalar multiplication and summation, using equations (7) and (8). This is a delicate process, as we require linear combinations of the function ϕ and a number of translated replicas to combine to produce the same profile, but dilated. It is a bit like a self-similar jigsaw puzzle: the jigsaw pieces at one scale have to fit together perfectly to produce the same jigsaw piece doubled in size.

4.2 Haar Wavelets


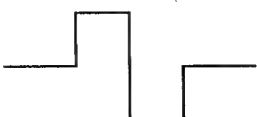
We demonstrate the use of the scaling and wavelet equations with the **Haar** scaling function and wavelet, which until the 1980's was the only example of a function of compact support known to solve these equations. We begin with the very crude lowpass and highpass filters $H_0 = [1/2, 1/2]$ and $H_1 = [1/2, -1/2]$. We now have the following dilation and wavelet equations:

$$\begin{aligned} \phi(x) &= \phi(2x) + \phi(2x-1) \\ w(x) &= \phi(2x) - \phi(2x-1) \end{aligned} \quad (11)$$

which have the Haar scaling function and wavelet as solutions, shown in Table 3.

In this simple case it is obvious how the Haar scaling function (top hat) solves the dilation equation, as the

Table 3. Haar scaling function and wavelet

Name	Function	Shape
Haar scaling function	$\phi(x) = \begin{cases} 1 & 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$	
Haar wavelet	$\phi(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$	

summands on the right hand side do not overlap, but the existence of other, less trivial solutions (with different filters H_0 and H_1) which are smooth and do overlap is much more interesting and useful. The Haar wavelet has some good properties; it is very compact, has a very simple two-point filter and its pyramid translations and dilations form an orthogonal basis. It is not smooth, however, and thus has very poor localisation of Fourier frequency. A major advance was made in the mid 1980's by Ingrid Daubechies, who discovered a family of smooth wavelets which also solve the dilation and wavelet equations. These are now widely used in signal and image compression.

4.3 Existence and Construction of Wavelets

The obvious remaining issue is knowing when there are solutions to the dilation and wavelet equations, are how to find them. Existence and uniqueness depend on the following Toeplitz matrix:

$$T(H_0) = \begin{pmatrix} h_0(1) & h_0(2) & h_0(3) & h_0(4) & \cdot \\ & h_0(1) & h_0(2) & h_0(3) & \cdot \\ & & h_0(1) & h_0(2) & \cdot \\ & & & h_0(1) & \cdot \\ & & & & \ddots \end{pmatrix} \quad (12)$$

derived from the lowpass filter H_0 . We derive another matrix, called the transition matrix:

$$T_0 = 2T(\downarrow 2)H_0 T(H_0) \quad (13)$$

The dilation and wavelet equations have a unique solution if the eigenvalues of the transition matrix are less than unity, except for a single eigenvector with unit eigenvalue. Moreover, when this happens we have a simple recursive recipe for calculating the scaling function ϕ ; it is the limit of the following convergent sequences of functions:

$$\phi_{n+1}(x) = \sum_k 2h_0(k) \phi_n(2x-k) \quad (14)$$

whose resemblance to the dilation equation (7) is obvious. The wavelet function can be derived directly from the scaling function using equation (8).

To summarise, we have achieved a huge gain in efficiency by calculating the wavelet transform using a discrete multiresolution filter bank, but at a price, as we have imposed a constraint on the wavelet function w in the form of the dilation and wavelet equations. For many applications the shape of the wavelet is not critical, as long as it has the required compactness in space or frequency, but there are some applications, for example target detection, where the shape is more important. We have also constrained the evaluation of the wavelet

transform to a discrete pyramid grid, which is also be unsuitable for applications where translation and scale invariance are important.

5. Wavelet Varieties

As with filter design, there are various, sometimes conflicting requirements of wavelet analysis, so there are different types of wavelets which are suitable for different applications, discussed in this section.

Although the scaling and wavelet functions are uniquely determined by Equations (7) and (8), they can still be controlled by the coefficients of the filter H_0 . The typical approach to wavelet design, therefore, is to design this filter first, along with H_1 . The reconstruction filter is then derived from the perfect reconstruction equations (6), which provides the inverse to the discrete wavelet transform. In this section we review briefly some of the many varieties of wavelets and filter banks that are available for 1D signal analysis. Higher-dimensional signals, including images, are considered in Section 6.

5.1 Orthonormal Wavelets

The most well known type of wavelet are the **orthonormal wavelets** discovered by Ingrid Daubechies. Here the filter H_0 is designed such that the analysing wavelet and all its translations and dilations on the pyramid grid are mutually orthogonal and have unit energy:

$$\int_{\mathbb{R}^1} w(2^l x + k) w(2^q x + p) dx = \begin{cases} 1 & k = p, l = q \\ 0 & \text{otherwise} \end{cases} ; \quad k, l, p, q \text{ integers} \quad (15)$$

Orthonormal functions are liked by mathematicians because transforms which use these functions are very stable, and trivial to invert, so reconstruction of the signal or image is very easy and efficient. In the case of the wavelet transform the inverse is given by:

$$f(x) = \sum_{p, q < q_0} b(p, q) w(2^q x + p) + \sum_p a(p, q_0) w(2^{q_0} x + p) \quad (16)$$

so the wavelet coefficients on the pyramid are the weighting factors required to reconstruct the signal or image f . A necessary and sufficient condition for orthonormal wavelets is that the filter H_0 is **double-shift orthogonal**, which means that when convolved with its transpose, all the even coefficients are zero except at zero, where the coefficient is two. The odd coefficients do not affect orthogonality of the wavelets. In the Fourier domain these filters are known as **half-band**, because the power spectrum added to a mirror image about half the Nyquist adds to unity at all frequencies:

$$|\hat{H}(\omega)|^2 + |\hat{H}(\omega + \pi)|^2 = 1 \quad (17)$$

The highpass filter H_1 is thus derived from H_0 by changing the signs of the odd coefficients and then transposing. The synthesis part of the filter bank is identical to the analysis part except for a transpose: $h_0(k) = f_0(-k)$ and $h_1(k) = f_1(-k)$.

The remaining task is to design the coefficients of H_0 to satisfy Equation (17). This is a complicated process, so only an outline of one method to derive orthogonal wavelets will be given here. First a power spectrum $|\hat{H}(\omega)|^2$ is found satisfying Equation (17), which for FIR filters means finding a finite symmetric polynomial satisfying:

$$P(x) + P(1-x) = 1 \quad (18)$$

but where for smooth, band-limited wavelets it is also desirable to have P and as many derivatives as possible zero at $x=0$ and $x=1$, except $P(0)=1$. The family of solutions, called maxflat filters, is given by:

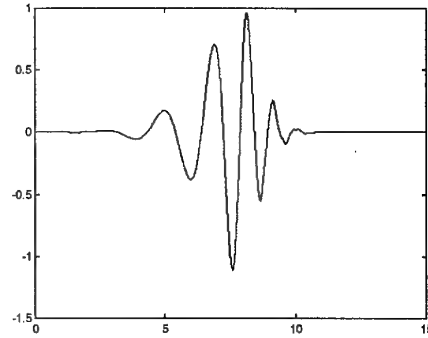
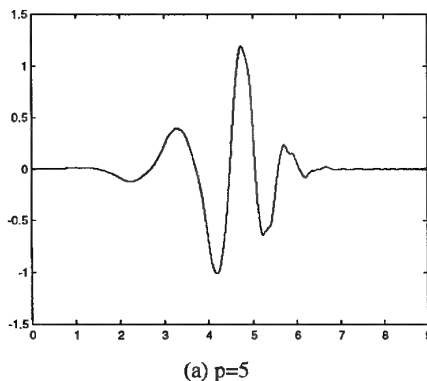
$$P(x) = (1-x)^p \sum_{k=0}^{p-1} \binom{p+k-1}{k} x^k \quad (19)$$

Next the coefficients of H_0 are derived from P ; P is the autocorrelation of H_0 :

$$P(z) = H_0(z)H_0(z^{-1}) \quad (20)$$

and solving this equation is known as **spectral factorisation**. One method is to find all the complex roots of P , which because it is real and symmetric, has roots which come in pairs which are mutually reciprocal. The polynomial H_0 is derived by gathering together one root from each pair whose modulus is less than or equal to unity. Fig. 3 shows the Daubechies' wavelets with $p=5$ and $p=8$, which become smoother and more band-limited with higher p .

Fig. 3. Daubechies' wavelets DB5 and DB8



Orthonormal wavelets and filter banks are very convenient, but the constraint imposed by Equation (17) is very restrictive. For example, except for the trivial case of the Haar scaling function, none of the scaling functions, wavelets or filters are symmetrical. It is tempting to use orthogonal wavelets because of their simple inversion formula, but in many cases this is unnecessary, as we often do not require the same coefficients for the analysis and synthesis filters. An analogue is the use of matrices to solve simultaneous linear equations. A matrix with a simple, sparse inverse permits us to solve simultaneous equations easily, but efficient inversion does not require the additional constraint of the inverse being equal to the transpose, as required of orthogonal matrices.

5.2 Biorthogonal and Semi-orthogonal Wavelets

Orthonormal wavelets are the analogue of orthogonal matrices. Likewise **biorthogonal wavelets** are the analogue of invertible matrices. The inverse of the filter bank is perfect reconstruction, so we still require Equations (6) to be solved, but now the synthesis filters F_0 and F_1 can be very different to the analysis filters H_0 and H_1 . We also have to work with two types of scaling and wavelet function: one pair for analysis, to calculate the wavelet coefficients using Equation (2), and a different pair for synthesis, to reconstruct the signal or image, using Equation (16). The wavelets and filter banks are still related by the dilation and wavelet equations (7,8), but now the analysis functions are generated by the analysis filters H_0 and H_1 , and the synthesis wavelets are generated by the synthesis filters F_0 and F_1 . The perfect reconstruction equations (6) ensure that these wavelets are biorthogonal, which means that in Equation (15) one of the wavelets in the integrand is an analysis wavelet, and the other is a synthesis wavelet, but otherwise the formula is the same.

Semi-orthogonal wavelets are another important variety, where wavelets of different scales are orthogonal, but wavelets of different position are not always orthogonal. These are useful for **interpolation** and **approximation** of functions. A popular simple choice are the **spline** wavelets, whose scaling functions

are the Haar scaling function (top hat) convolved with itself n times, and whose lowpass filter has binomial coefficients. The orthogonality across scale ensures that the accuracy of approximation for smooth functions increases with maximum rapidity as scale decreases, but orthogonality between wavelets of the same scale imposes undesirable constraints which degrade approximation

A useful tool which has gained a lot of attention recently is **lifting**, which is a systematic and flexible method of constructing biorthogonal wavelets and filter banks. The idea is to change H_0 to meet application-specific design requirements, whilst still satisfying the perfect reconstruction of Equation (6). It turns out that any change to H_0 of the following form will achieve this:

$$H_0(z) \rightarrow H_0(z) + F_0(-z)S(z^2) \quad (21)$$

for any filter $S(z)$. We can do a similar operation on the synthesis filter F_0 , which is called **dual lifting**. Typically the process of filter design starts with a simple filter, for example a delta function or top hat, called a "**Lazy filter**", and then the processes of lifting and dual lifting are iterated with suitable choices of S , until the design requirements are met.

5.3 Wavelet Frames

Lastly, we briefly mention **wavelet frames**. The discrete wavelet transform and filter banks mentioned so far are fully invertible transforms, so there is a one-to-one correspondence between the signal and the output of the wavelet transform or filter bank. This is equivalent to the translations and dilations of the mother wavelet on the pyramid grid being a basis; they are linearly independent and span the space of signals. In wavelet frames the requirement for independence is dropped, which typically involves oversampling the continuous wavelet transform by adding extra points to the pyramid, for example by doubling the resolution in position or by halving scales between octaves. The wavelets still span the signal space, so any signal can be recovered from the wavelet transform. Not all such functions of position and scale are wavelet transforms, however, so wavelet frame transforms only have one-sided inverses.

Wavelet frames are generally more computationally intensive, as there are additional coefficients to calculate, but objects such as targets can be characterised more flexibly at intermediate positions and scales. Wavelet frames become more translation-invariant as the sampling density increases, as they are better approximations to the continuous wavelet transform.

To summarise, there is a wide variety of wavelets and filter banks available for signal analysis, each with its own strengths and weaknesses. Although it is tempting to use the first family of wavelets that springs to mind,

for example the popular Daubechies wavelets, there may be others more suitable for the application. There are also design techniques, such as lifting, to customise wavelets, should off-the-shelf varieties not suffice.

6. Wavelet Analysis in Higher Dimensions

The techniques described in Sections 3–5 are applicable to 1-dimensional signals. In higher dimensions there are two approaches to wavelet analysis: either to use **separable filters** which can be derived easily from 1D filters using exterior products, or **non-separable filters**, which have to be designed from scratch, which is more difficult.

6.1 Separable Wavelets

Separable functions of several variables are **Cartesian products** of functions of fewer variables:

$$f(x_1, x_2, \dots, x_n) = f_1(x_1)f_2(x_2)\dots f_n(x_n) \quad (22)$$

where in general the arguments x_k can be vectors as well as scalars. Exterior products of scaling functions and wavelets make effective higher-dimensional wavelets, inheriting all the properties of their lower-dimensional components. To simplify the notation, we will consider exterior products of two 1-dimensional wavelets to facilitate image wavelet analysis, but the principles behind higher-dimensional wavelet analysis are identical.

Image wavelet analysis involves one lowpass filter H_0 and *three* highpass filters, H_1, H_2, H_3 , each of which is the exterior product of 1-dimensional lowpass or highpass filters:

$$\begin{aligned} h_0(m, n) &= h_0(m)h_0(n) & h_1(m, n) &= h_0(m)h_1(n) \\ h_2(m, n) &= h_1(m)h_0(n) & h_3(m, n) &= h_1(m)h_1(n) \end{aligned} \quad (23)$$

Similarly there is one scaling function and *three* wavelet functions formed as exterior products of their 1-dimensional counterparts:

$$\begin{aligned} \phi(x, y) &= \phi(x)\phi(y) & \phi_1(x, y) &= \phi(x)\psi(y) \\ \phi_2(x, y) &= \psi(x)\phi(y) & \phi_3(x, y) &= \psi(x)\psi(y) \end{aligned} \quad (24)$$

The multiresolution filter bank has four outputs at each scale; the lowpass output is downsampled and goes to the next resolution, and the other 3 outputs are the detail or wavelet coefficients, as for the 1-dimensional case. The three types of wavelet are usually regarded as having horizontal, vertical and diagonal orientation.

The discrete wavelet transform is usually displayed as shown in Fig. 4, though this representation can be misleading. In this representation the density of wavelet

coefficients is kept constant, with larger regions required to store information at high resolution (low scale). The wavelet coefficients at any scale are three times the number at all larger scales, because there are three highpass filters to one lowpass filter. The regions are designated LL (lowest resolution only) HL, LH and HH according to which combination of 1-dimensional filters is used in the Cartesian product. This representation is convenient, because the transform has the same shape and number of pixels as the original image, an example shown in Fig. 5, but the larger scales are portrayed as being smaller in size! It is true that the downsampling operator has this effect, but a more natural interpretation is that the wavelet filters increase in scale.

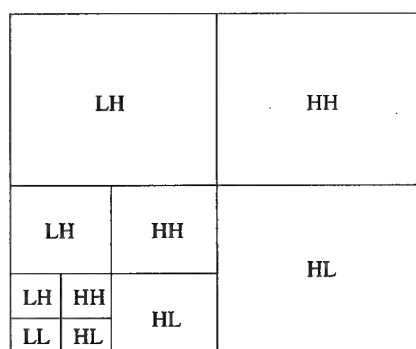


Fig. 4. Image wavelet display

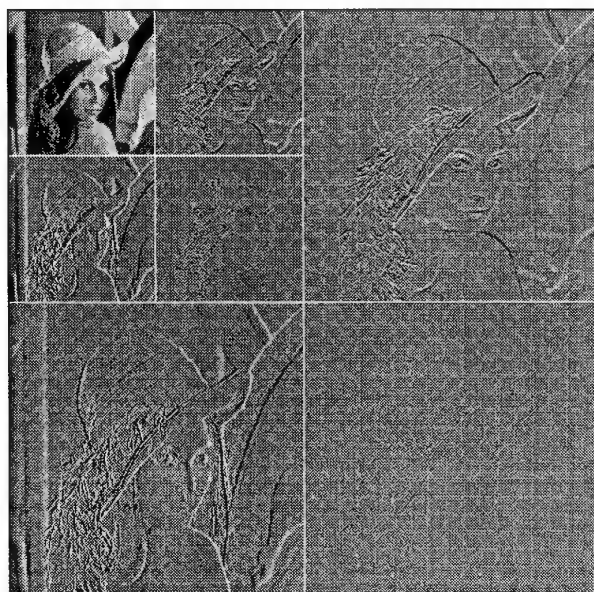


Fig. 5. Wavelet decomposition of 'Lena' - 2 levels

6.2 Non-Separable Wavelets

The alternative approach to image wavelet analysis is to use non-separable wavelets. Although more difficult to design, these can be more flexible, especially in orientation. The image pyramid grids and resampling do not need to be rectangular or separable, either. An example is given in Fig. 6, where the small and large dots comprise the grid at one resolution, and the large dots only comprise a sub-grid at the next lowest resolution.

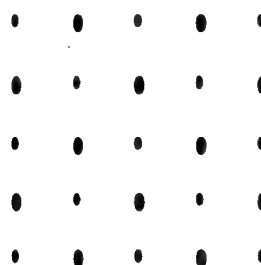


Fig. 6. Non-separable grid

In this example the change in area and the resampling factor between scales is not 4 as it would be in the separable case, but 2, so there is only one highpass filter required, as for the 1-dimensional signal case. In this case resampling causes a rotation through 45° . This is known as **quincunx** resampling. Hexagonal grids can also be used, which permits wavelets with 60° orientation intervals to be constructed. Even more exotic wavelet grids have become popular in the interpolation of complex geometric surfaces [7], which is a very active research topic.

7. Wavelet Packets

In conventional wavelet analysis the main source of variety in the transform is in translation and dilation. One or a very small number of filters is involved, except for differences in position and scale. This limits the variety of information that individual wavelet coefficients represent. Another approach that has gained popularity in recent years is that of **wavelet packets**, where the functions used to represent the signal or image *vary in shape* also. Typically frames are used instead of bases, initially providing redundancy, but then a subset of the coefficients are selected to derive a basis which is adapted to fit the incoming data.

One way to do this is to extend the sub-band coding to encompass any **dyadic tree** structure. In conventional wavelet analysis it is only the lowpass filter that is split further by downsampling and bandpass filtering; the

output from the highpass filter is left alone. In a more general dyadic tree, the decision to split the channel is applied more arbitrarily, to yield a wide variety of transforms. Fig. 7 shows some examples of dyadic trees.

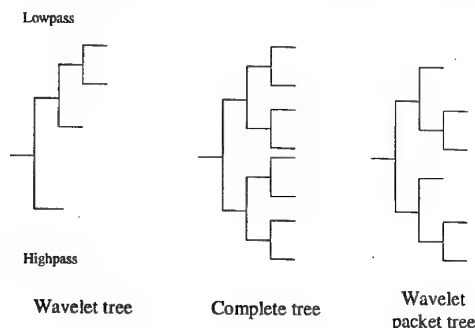


Fig. 7. Dyadic tree structures

The complete dyadic tree divides all branches, resulting in an equal partition in the Fourier domain, analogous to the short-time Fourier transform which divides the signal into a set of time-frequency cells of identical duration and frequency bandwidth. If we apply the complete tree to the Haar filter, for example, we get the **Walsh functions**, shown in Fig. 8.

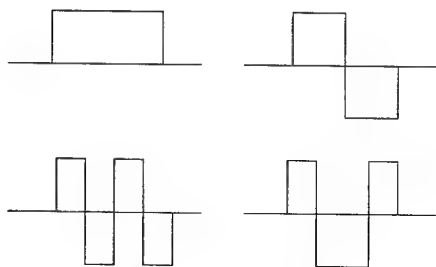


Fig. 8. Walsh functions

In general the aim of wavelet packet analysis is to approximate the signal or image by a series of functions chosen from a large set, called a **dictionary**, for example the functions generated by all dyadic trees. The functions are chosen to give the best approximation with the smallest number of components. The larger the dictionary, the more computation required, but also the greater the potential for an efficient representation. An alternative approach is to extend the transformations which generate the wavelet basis beyond translation and dilation to include shape changes, for example frequency shifts and chirp angles (chirplets [8]), or in the case of image wavelets, affine transformations (ridgelets [9]).

There are also different approaches to choosing the functions from the dictionary to approximate the signal or image. One method is the **best basis** algorithm [10] which selects functions from a union of several bases. Another is **matching pursuit** [11], where wavelets are

selected from a large dictionary (e.g. generated by translation, dilation and frequency shifts) in the order that most rapidly decreases the approximation error, and at each stage subtracts the chosen function from the signal or image. Another method [12] is selection from a continuum of functions analogous to the continuous wavelet transform, searching for local maxima in correlation with the signal or image, but where parameters are not limited to position and scale, or to a discrete grid. A conjugate gradient search is used to refine the wavelet parameters after an initial grid search, enabling the wavelets to fit the signal or image data more accurately, and achieving invariance with respect to translation, dilation and related operations.

9. References

1. G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Rev. Ed., 1997.
2. Y. Meyer, *Wavelets, Algorithms and Applications*, Siam, Philadelphia, 1993.
3. Feder J., *Fractals*, Plenum Press, 1988.
4. A.K.Jain, "Image Data Compression: A Review", *Proc. IEEE*, **69**, pp.349-389, 1981.
5. T.H. Koornwinder (ed), *Wavelets: An Elementary Treatment of Theory and Applications*, World Scientific, 1993.
6. G.H.Watson, "The Detection of Unusual Events in Cluttered Natural Backgrounds", *NATO RTA lecture series 216, Application of Mathematical Signal Processing Techniques to Mission Systems*, 1999.
7. A.W.F.Lee et al, "MAPS: Multiresolution Adaptive Parameterisation of Surfaces", *Computer Graphics Proceedings (SIGGRAPH 98)*, pp.95-104, 1998.
8. S.Mann and S.Haykin, "The Chirplet Transform: Physical Considerations", *IEEE Trans. on Signal Processing*, **43**(11), Nov 1995.
9. E.J.Candes, "Ridgelets: Theory and Applications", PhD Thesis, Dept of Statistics, Stamford University, 1998.
10. R.R.Coifman and M.V.Wickerhauser, "Entropy-Based Algorithms for Best Basis Selection", *IEEE Trans. on Information Theory*, **38**, pp.713-8, 1992.
11. S.G.Mallat, "A Theory of Multi-Resolution Signal Decomposition: Wavelet Decomposition", *IEEE PAMI*, Vol. 1, pp. 674-693, 1989.
12. G.H.Watson and K.Gilholm, "Signal and image feature extraction from local maxima of generalised correlation", *Pattern Recognition* **31**(11) pp.1733-1745, Nov 1998.

© British Crown copyright 1999. Published with the permission of the Defence Evaluation and Research Agency on behalf of the Controller of HMSO.

The Detection of Unusual Events in Cluttered Natural Backgrounds

G.H.Watson

Room 1052, A2 Building, DERA Farnborough, Ively Road, Farnborough, Hants, GU14 0LX, UK

1. Introduction

This paper is concerned with the use of wavelet analysis and statistical models of natural backgrounds as a means of detecting unusual events within, in particular targets of military interest. The underlying principle is to detect targets as objects that stand out from the background, and hence are unusual, rather than searching for objects with prescribed characteristics and dealing with clutter as an afterthought. First a method of feature extraction is described based on wavelet analysis which is used to characterise both backgrounds and unusual events. Then the statistics of these features for natural backgrounds are considered, making use of fractal geometry, from which basic clutter rejection can be implemented. More advanced clutter rejection methods are then considered, based on the multivariate statistics of additional measurements. Three cases are considered in detail: the wavelet analysis of multispectral data, the use of local variance to reject clutter in intermittent backgrounds, and the use of temporal variability to reject clutter in image sequences.

The approach of modelling the background, rather than the target, has the advantage that little or no prior knowledge of the latter is required, leading to greater flexibility and robustness. Target prior can be added at a later stage, if available, for further discrimination and clutter rejection. In some military circumstances early warning of targets is required before any detailed structure can be resolved, the limiting case being point targets with a single-pixel signature. In such cases target prior is of little use in recognition, being limited to the time signature of a single pixel, so the use of background context can be critical to early detection.

The method to be described comprises the following stages:

- (a) Decompose the signal or image data into a set of discrete features which are suitable as an ensemble for characterising both targets and the background. These features are generally simple geometric shapes to facilitate their extraction, such as blobs and bars in images, but which can be combined to characterise more complex objects, such as roads and cloud edges. These features are usually extracted at multiple scales, using wavelet analysis.
- (b) Construct a statistical model of the background based on the above feature decomposition. Most natural backgrounds are difficult to model, having strongly non-Gaussian statistics and phase correlation, for example in the form of strong edges. In general the joint statistics of feature parameters such as brightness, position, scale and orientation need to be calibrated, resulting in multidimensional probability distributions. However, most natural backgrounds are stationary and exhibit fractal geometry, which simplifies the statistical modelling.
- (c) Extract potential targets as statistical outliers, that is at the edges or tails of the background distribution. Each object is assigned a prior probability that it belongs to the background, and hence not a target.
- (d) If additional target prior is available, use Bayes' formula to combine the prior distributions of targets and the background to estimate the a posteriori probability (likelihood) of there being a target. This topic is not covered in this paper.

This method can be applied to a wide variety of data, including 1-dimensional signals (e.g. acoustic data), 2-dimensional images (including multispectral imagery), 3-dimensional images (e.g. medical tomography), and time-sequenced imagery, where movement is part of the feature characterisation. The method is only limited by the methods of feature extraction available, and the accuracy of the background statistical models.

The remainder of this paper is organised as follows. Section 2 describes methods of feature extraction, based on searching for local extrema in the wavelet transform and analogous correlation, and explains the relationship between this and matched filtering. Section 3 describes how the statistics of these wavelet features are calibrated with the aid of fractal geometry. Section 4 explains how improved clutter rejection can be implemented by introducing additional random variables, and gives three examples: multispectral imagery, strongly intermittent backgrounds, and image sequence clutter rejection based on space-time filters.

2. Feature Extraction

As explained in Section 1, the purpose of feature extraction is to decompose the signal or image into a set

of discrete geometric components which are sufficient to characterise both targets and background sufficiently well that the former can be recognised as unusual events. This involves making measurements of the data for which targets might have unusual values, which in signal processing parlance means applying filters to the data. The most well understood filters are linear, which will be considered in this section; Sections 4.2 and 4.3 give examples of non-linear filters which provide further discrimination of targets and clutter.

2.1 Features from Matched Filters

Matched filter theory [1,2] can be used to derive the optimum linear filter (matched filter) to detect any prescribed object, in the sense that signal to noise ratio (SNR) is maximised. If f is the filter, x is the target then SNR is defined to be:

$$\frac{|f(x)|}{\sqrt{E(f(u)^2)}} \quad (1)$$

where E denotes expectation. In this section it will also be assumed that the signal or image data is stationary, that is statistically translation-invariant. This is usually the case for time-varying signals and imagery projected as a plan view, but there is often a statistical dependence on the vertical image co-ordinate for forward-looking imagery. The latter situation is considered in Section 4.3. For translation-invariant data the filters should be translation-invariant to avoid statistical bias, which implies linear filtering is equivalent to convolution. In such cases the Fourier transform of the matched filter $F(\omega)$ is given by:

$$F(\omega) = \frac{\bar{X}(\omega)}{N(\omega)} \quad (2)$$

where X is the Fourier transform of the target, the bar denotes complex conjugation, and N is the power spectral density (PSD) of the background. The inverse Fourier transform can be used to derive the convolution kernel in the signal or image space.

Matched filters are simple and effective when the target configuration is simple, for example a point target, as the number of possible target configurations may be small. Where the target is more complex, either because spatial structure can be resolved or because its trajectory is varied, the number or complexity of matched filters makes their implementation more difficult. For example, an aircraft may be viewed from many ranges and aspects, each requiring a different matched filter. In this paper matched filters are designed instead to detect simple geometric structures which are suitable for characterising parts of targets or backgrounds. For example, if the target is a missile, we may choose a bar shaped

component; this will not fit the missile exactly, but gives a fairly good approximation to the missile body and the plume. If we choose very simple, generic components, we will be able to pick up a wide variety of objects, but there will be some loss in detection sensitivity when the SNR is very low.

2.2 Translation and Dilation Invariance

Convolution involves correlating the kernel with the data and repeating this operation over all translations of the kernel. Translation is usually required because there is little or no knowledge of where the target is located in space or time prior to its detection. It is fortunate that many backgrounds are also statistically translation-invariant, otherwise translation of the filter kernel would not give consistent answers. Non-stationary backgrounds, such as in forward-looking imagery are a common source of false alarms for this reason.

In many situations the scale of the target signature is also not known *a priori*, primarily because it is affected by the distance to the sensor. Thus multiresolution analysis is required, where the filtering is repeated over varying scale as well as varying position. This is why wavelet analysis is a useful tool in target recognition. Even when there is prior knowledge of target size, the geometric features used in its analysis may be of varying scale, for example a missile's guidance fins are usually much smaller than its fuselage.

Many backgrounds are statistically scale-invariant, though this is not as intuitively obvious as translation-invariance. In many images of familiar scenes, such as natural terrain, the scale or magnification is hard to discern unless there is a reference object of known size, such as an adult human. This phenomenon is the subject of fractal geometry [3,4], where scale-invariance is known as self-similarity. Strictly speaking self-similarity means that there is a scaling transformation under which the signal or image is identical to a subset of itself, where the scaling transformation is given by:

$$f\left(\frac{x}{s}\right) = s^{-h} f(x) \quad (3)$$

where f is the signal or image being transformed, s is the scale or dilation factor and the constant h is sometimes called the self-similarity parameter. When modelling natural backgrounds the equality of equation (3) is replaced by statistical invariance. The response of statistically self-similar backgrounds to filters at different scales can be normalised by dividing by s^h [5], allowing the same threshold to result in the same false alarm rate over a range of scales. We will return to the use of fractal geometry in modelling the statistics of backgrounds in Section 3.

2.3 The Wavelet Transform

So far we have established that correlation should be combined with translation and dilation, resulting in wavelet analysis, and that the background is often translation and scale invariant, which means that the statistics of the wavelet transform should be uniform. However, care must be taken to employ the correct scale normalisation factor s^{-h} , to ensure uniformity. The standard definition of the continuous wavelet transform is consistent with $h=-1/2$, the self-similarity parameter for uncorrelated backgrounds (white noise):

$$T(p, s) = s^{-1/2} \int_{-\infty}^{\infty} f(x) g\left(\frac{x-p}{s}\right) dx \quad (4)$$

but in general, for constant false alarm rate (CFAR) detection, the formula should be modified to:

$$T(p, s) = s^{-1-h} \int_{-\infty}^{\infty} f(x) g\left(\frac{x-p}{s}\right) dx \quad (5)$$

The shape of the analysing wavelet g can be derived as the matched filter of the geometric feature which will be used to characterise the signal or image data, at a chosen position and scale. If more than one type of feature is used, for example radial basis functions and oriented bars, then more than one analysing wavelet is required, and more than one wavelet transform calculated.

The use of matched filter theory to derive appropriate analysing wavelets shows that the latter depend on both the shape of the geometric features used to characterise the data, and on the properties of the background, in particular its PSD. For uncorrelated backgrounds the matched filter is the same as the geometric feature, except for a reversal in space and time (a mirror image). For most correlated backgrounds the PSD decreases with frequency, so the matched filter is similar to the feature, except higher frequencies are emphasised, which has the effect of introducing side-lobes. For self-similar backgrounds the PSD is a (usually negative) power of frequency [5], and the matched filter is thus a fractional derivative [6] of the feature. For example, in Brownian noise the matched filter is the time-reversed second derivative of the feature, and the equivalent operator for rotationally-symmetric image backgrounds is the Laplacian operator. In the latter case the matched filter for a Gaussian radial basis function is its second derivative, which is similar to the Difference of Gaussian (DoG) filter so popular in target detection.

2.4 Local Maxima

The final stage in feature extraction is searching for local maxima in the absolute value of the wavelet transform, as this enables the data to be decomposed into a discrete set of objects, and reduces redundancy. Restricting

measurements to local maxima, rather than at all points above a threshold, prevents redundant features being generated from the same object in its neighbourhood. At each local maximum, its location in wavelet space (position and scale), and the type of analysing wavelet (e.g. matched to a blob or bar) are recorded, as well as the wavelet transform value itself. These values provide a lot of information about targets and backgrounds, for example the location, scale, orientation and grey-level contrast of edges and bars, and an approximate reconstruction is available by linear superposition of the features whose parameters match the local maxima. In the context of target detection, it is more useful to retain only those features whose amplitude (wavelet transform value) exceed a threshold corresponding to a prescribed probability (Section 3). These features provide concise information about the target and effective clutter rejection, from which a partial reconstruction of the target can be obtained. Examples of feature extraction and partial reconstruction are given in Fig. 1, demonstrating the ability to represent artefacts such as roads and buildings, whilst rejecting most clutter.

The method of searching for local maxima depends on the application. A search restricted to the positions and scales of the discrete wavelet transform is quick to implement, but the resulting features are limited by the poor resolution of the pyramid grid of positions and scales. This type of feature extraction is also not truly translation and scale invariant, because most shifts in position and scale cause the wavelet grid to change. This drawback is mitigated if a wavelet frame is used with a higher resolution in position and scale, which is equivalent to interpolating the wavelet transform. Better still, but more costly, is to refine the positions and scales of the features thus found by evaluating the continuous wavelet transform explicitly using expanded filters instead of downsampling the signal or image, and then optimising this function using a local search, for example a conjugate gradient search [7]. This method results in true translation and scale invariance.

3. Background Statistics and Fractal Geometry

3.1 Threshold Exceedance Model

Once geometric features have been extracted from the data (Section 2), the next stage is to discriminate between potential targets and the background, based on the statistics of the latter. First we describe a simple statistical model for backgrounds with fractal geometry which predicts the rate at which feature amplitude (wavelet transform value) exceeds any prescribed threshold y as a function of scale s . Jones [8,9,10] has shown that amplitude threshold exceedance rates are proportional to a power of scale, s^{-D} , where D is a fractal dimension describing the self-similarity of feature

population, which is different to h (Section 2), describing the self-similarity of feature amplitude. This self-similarity enables a single scale-independent distribution to represent the joint statistics of amplitude and scale:

$$N_y s^D = F(y) \quad (6)$$

where N_y is the threshold exceedance rate (number of features whose amplitude exceeds the threshold y), and F is a scale-independent function defining the shape of the distribution.

The threshold exceedance statistics in [8] and for the data analysed in this paper are approximated well by the generalised exponential function $F(y) = y^c/\beta$ where c and β are constants. This leads to the following model for threshold exceedance statistics:

$$\log\left(\frac{N_y s^D}{\alpha}\right) = -\left(\frac{y}{\beta}\right)^c \quad (7)$$

where:

- N_y is the number of local extrema at scale s whose amplitude $T(p,s)$ exceeds the threshold value y . N_y as a function of y is closely related to the cumulative probability distribution of wavelet amplitude T at the local extrema.
- D is a *fractal dimension*, and represents the scaling behaviour of the population density of local extrema with respect to position. For statistically stationary signals the fractal dimension is equal to the topological dimension, that is 1 for signals, 2 for images and 3 for image sequences or tomography. For intermittent self-similar data the fractal dimension can be less than the topological dimension because of the hierarchical clustering of local extrema, where at any scale a cluster is formed of several smaller clusters with intervening gaps. The Cantor set is an example of a 1D signal with a fractal dimension of less than unity.
- α is called *lacunarity*, which is a measure of overall population density of local extrema in a fractal.
- β is a measure of overall intensity, which in a stationary process would be proportional to the standard deviation of the signal or image.
- c is an inverse measure of the strength of the tails in the distribution of wavelet amplitude. For example $c=1$ represents an exponential distribution with moderately strong tails,

whereas $c=2$ represents a Gaussian distribution with weak tails.

Equation (7) can be used to assign an *a priori* probability that a geometric feature arises from the background, and by setting a threshold on this probability, driven by the highest acceptable false alarm rate, the detection of targets as unusual events is possible. For backgrounds with translation and rotation invariance, we thus have a simple method of target detection with no bias in position, scale or orientation.

3.2 Calibration of the Model

To use equation (7) we need to know the values of the parameters D , α , β , c and the self-similarity parameter h (Section 2). In some cases there is prior knowledge of some or all of these parameters, for example there are theoretical reasons for assuming a self-similarity parameter of $h=1/3$ for fully-developed turbulent flow [11]. More often it is better to estimate these parameters from the signal or image data. This can be done off-line as a separate training process, with the advantage that there will probably be independent assurance that targets are missing from the data. On-line calibration has the advantage that the resulting detection algorithm can adapt to changing background conditions, for example caused by changes to the weather.

We now describe a simple and efficient method of estimating D , α , β , c and h , making frequent use of linear least-squares regression. The data used in this statistical estimation is assumed to be a set of feature amplitudes and scales (y_k, s_k) . First, an approximate value of self-similarity h_0 is used to calculate the modified wavelet transform of equation (5). This can be estimated with reasonable accuracy from the slope of the PSD plotted in log-log co-ordinates; in theory the PSD should be proportional to $s^{-(2h+1)}$ [5]. The self-similarity parameter will be refined based on estimation from feature amplitudes and scales. The initial guess only weakly affects the location and density of local extrema in the wavelet transform, so high accuracy is not required.

Next, threshold exceedance counts N_y are derived, conditioned by scale, for a range of scales s and amplitudes y . The conditioning of scale depends on the method of searching for local extrema in the wavelet transform. If the search is limited to values of scale on a discrete grid, then these values are used to partition the data, and threshold exceedance counts are computed for each partition set. On the other hand, if scales are refined, for example using a gradient search, it is necessary to partition scales into a set of intervals. Threshold exceedance counts N_y are derived for a range of values y for each scale or scale interval. An efficient way of doing this for every value of y in the geometric feature set is to sort the amplitudes of the latter into

descending order; the value N_y for each y is then its index in the sorted array.

The next step is to assign a weight to each pair of (y, N_y) which reflects its relative importance when fitting the model given by equation (7). The weights depend on the confidence in the data at the tails of the distribution where y is large and N_y is small. Where we are sure that there are no targets, i.e. the background is "pure", it is appropriate to assign relatively large weights at the tails, as it is the latter which are important in discriminating targets from clutter. In this case assigning weights which have uniform density in amplitude y , or even density which increases with y , is appropriate, which can be done by setting each weight to be the difference between y and the next lowest value of y in the same scale partition. Alternatively, where there may be targets present, for example during on-line calibration, it is better to give the tails low weighting, and rely on the model in equation (7) to extrapolate these. This can be done by assigning equal weights to each pair (y, N_y) , as the number of target features is almost always orders of magnitude lower than the number of background features.

The parameters D, α, β, c and h , are now estimated using linear least-squares regression with the weights described above. First we need to rewrite equation (7) to allow for the difference between the approximate self-similarity parameter h_0 and the actual self-similarity h :

$$\log\left(\frac{N_y s^D}{\alpha}\right) = -\left(\frac{y}{s^{h-h_0}}\beta\right)^c. \quad (8)$$

The tail strength c is estimated by fitting a linear model to $(\log(y), \log(\log(N_y)))$ and measuring the slope. Next a linear model is fitted to $(-y^c, \log(N_y))$ for each scale partition, as follows:

$$\log(N_y) = A_s - B_s y^c \quad (9)$$

where the slopes and intercepts A_s and B_s depend on scale s . Then linear models are fitted to the following data: $(\log(s), A_s)$ and $(\log(s), -\log(B_s))$ resulting in:

$$A_s = P_1 + P_2 \log(s) \quad (10)$$

and

$$-\log(B_s) = P_3 + P_4 \log(s) \quad (11)$$

Substituting (10) and (11) into (9) we obtain:

$$\log\left(\frac{N_y}{\exp(P_1)s^{P_2}}\right) = \frac{-y^c}{\exp(P_3)s^{P_4}} \quad (12)$$

Comparing equations (8) and (12), the parameters D, α, β , and h , are derived straightforwardly from the slopes and intercepts P_1, P_2, P_3 and P_4 .

4. Clutter Rejection from Joint Statistics

For some backgrounds, for example where clutter is intermittent, detection sensitivity can be increased by taking additional measurements into account, thus introducing additional criteria for discriminating targets and the background. In this section we describe a method of doing this based on the modelling of the joint statistics of wavelet amplitude and other variables describing the background.

4.1 Vector Fields

The most straightforward example of additional measurements is the analysis of vector fields rather than scalar data, in which each sample pixel has a vector value. Examples of vector fields are fluid flow velocity, electromagnetic fields, and multispectral imagery. In this case the joint statistics of wavelet amplitudes of each component of the vector (treated as a scalar field) are modelled. These wavelet amplitudes are the components of a multiresolution vector field which is a vector analogue of the wavelet transform:

$$\mathbf{T}(p, s) = s^{-1-h} \int_{-\infty}^{\infty} \mathbf{f}(x) g\left(\frac{x-p}{s}\right) dx \quad (13)$$

where correlation now involves pointwise multiplication of the vector field \mathbf{f} by the scalar filter g and integration is vector summation. Matched filter theory can still be used to design analysing wavelets, but now the power spectrum contains information about the cross correlation of the vector components at each frequency, as well as their autocorrelation.

There is no direct analogue of the feature extraction of Section 2 for vector fields, because local maxima of the wavelet transform are not defined, being vector valued. If the probability density function (PDF) of the wavelet transform were known a priori and easy to compute, then we could derive a scalar-valued function of the wavelet space by composing the above vector wavelet transform (which maps wavelets to a vector space) with the PDF (which maps the same vector space to the real line). It makes sense to search for local minima in the composite function, as these correspond to locally unusual events. There are two drawbacks to this:

- Joint PDFs are difficult to estimate accurately, with the difficulty increasing rapidly with the number of vector components. Over-fitting the PDF to the training data is a common problem [12].
- Not all outcomes of low PDF correspond to likely targets. For example, when based on local maxima, most distributions of wavelet amplitude have low PDF near the origin, corresponding to very low

feature brightness or contrast. These features are not usually target related.

In view of the above problems, the usual approach is to search for local maxima of a derived scalar quantity, the Mahalanobis distance [13]:

$$D(T) = (T - \mu)^T C^{-1} (T - \mu) \quad (14)$$

where μ is the vector mean of the wavelet transform values T , and C is the multivariate covariance matrix of T . For Gaussian distributions D is proportional to the negative logarithm of the PDF, and so is an unbiased measure of rarity. This measure is also only high for unusually high wavelet amplitudes (relative to the mean) at the external boundary of the distribution, so ignores very faint features. Like principal component analysis, this method is very effective at extracting features with unusual vector direction as well as amplitude, especially when there is a high level of correlation in vector direction. This situation is common in multispectral imagery when the spectral bands of the background are close in radiation frequency, resulting in a limited range of "colour". Objects which have an unusual radiation spectrum (colour) often generate wavelet amplitudes with a large Mahalanobis distance, even when none of the vector components is unusually large regarded in isolation.

The example in Fig. 2 concerns 5-band AVHRR (Advanced Very High Resolution Radiometer) satellite imagery of sea surface background. This data is a fusion of two visible bands (0.6 and 0.8 μm) and three infrared bands (3.7, 11 and 12 μm). The objects of interest are tracks which arise from ship plumes altering the spectral properties of the cloud. These tracks are visible in Band 3 (3.7 μm), and are generally absent or very faint in the other bands. Figs. 2a and 2b show an example of such a track against cloud clutter in Bands 1 and 3 respectively: the local SNR of the track is greatest in Band 3, but not sufficient to be easily distinguished from the cloud edges, especially towards the top of the picture. We base our analysis on these two bands because the other bands are either very similar in frequency or contain little new information.

Figs. 2c and 2d illustrate the advantage of basing the Mahalanobis distance on the wavelet transform T (Fig. 2c) rather than the raw pixel values (Fig. 2d); the ship track stands out much more clearly in the former case.

4.2 Intermittent Backgrounds

Another example of the use of additional random variables concerns the modelling of strongly intermittent backgrounds, in which some regions are energetic (highly cluttered) and others are relatively quiescent. Statistical translation-invariance does not rule out this

kind of variability with respect to position; it implies only that *global* statistics are invariant. *Local* statistics, such as standard deviation over a neighbourhood of the signal or image, may still be position-dependent, and often are for natural backgrounds. Variation in local statistics happens either because the image background comprises more than one texture, for example associated with woodland, mountains and sea, or because a single texture type is spatially intermittent, in which some regions are more energetic than others.

Intermittency is represented implicitly in the statistical model of Section 3 by the fractal dimension D and the lacunarity α , but no account is taken of the dependence of the average amplitude or population density of the geometric features on their position, i.e. their tendency to cluster. Thus whilst the effect of clustering on global population densities is accounted for, the variation in local density with position is not. If position dependence is not modelled, the same thresholds will be applied to energetic and quiescent regions alike, resulting in either a loss of sensitivity in the latter or too many false alarms in the former. We now show how the joint statistics of wavelet amplitude and a non-linear function, local energy, is used to rectify the above limitation. Further detail is given in [14].

The most obvious method of allowing for spatial variations in background activity is to condition the background model given by equation (7) with respect to position. This approach would require statistical distributions to be measured over neighbourhoods of the signal or image and then the model parameters fitted in each case separately. This approach is not effective because the fractal model parameter estimation only works well with large images (at least 64x64 and usually much larger) as filter statistics need to be compiled over a number of scales and in sufficiently large numbers to estimate a cumulative distribution function. The fractal model has the advantage of realism when applied to large images but at the cost of a rapid decrease in accuracy for images much below 128x128 pixels, and is thus unsuitable for the representation of local statistics.

In any model of the background which is conditioned by position, there is an inevitable conflict between the accuracy of the estimation of statistical parameters on the one hand, and spatial resolution of background intensity levels on the other. For the purposes of target detection, good spatial resolution is important because the threat, for example an incoming missile, may be very close to clutter, such as cloud edges, which could lead to incorrect thresholds being used by the detection process. The advantages of measuring background activity levels with spatial resolutions of less than 64 pixels in such cases are obvious. It follows that in the choice of the representation of local statistics, spatial resolution should be given a high priority. To achieve high spatial resolution, local background activity needs to be

represented by a robust statistic which requires a small amount of image data to compute reasonably accurately. This is why a simple non-linear measure, local energy, has been chosen for the spatial conditioning of the background statistics.

Local energy is a measure of local background activity which is computed over neighbourhoods of *varying size* surrounding each geometric feature extracted from the data (Section 2). Local energy $E(x, L)$ is defined to be the variance of the signal or image over a region centred on x and of size L . For 1D signals this region is a subset of the real line, for example an interval of length L , for 2D images it is a subset of the real plane, for example a square, and for 3D image sequences a subset of space-time. Detection sensitivity is generally improved if the support of the geometric feature is subtracted from this region, so that the wavelet amplitude and local energy are less mutually dependent.

The region size L is defined relative to the scale s of each extracted feature and is not to be confused with the scale of the feature itself. Local energy measurements over variable size L are required to achieve the best balance between spatial resolution and accuracy of local energy measurement. In cases where a potential target is close to but not within an energetic region, a small local energy region gives the best probability of detection. In cases where the background activity is more uniform, however, fainter targets can be detected with greater sensitivity where background levels have been measured over an extended region. The following region sizes L have been considered in this paper: 2, 4, 8, 16 and 32. We now describe how the joint statistics of local energy at each size L and wavelet amplitude T are used to model spatially intermittent backgrounds and improve the discrimination between targets and clutter, illustrating the technique on infrared imagery of the celestial background (Fig. 3) in which some simulated point targets have been inserted.

Fig. 4 shows the joint statistics of T and E in the form of a scatter plot, where L is set to 8, background data are presented by points, and the targets by crosses. The scatter plot resembles the joint PDF of these random variables, as regions of high population density in the scatter plot correspond to regions of high PDF. The embedded targets are mostly separate from the background distribution, that is in a region where the background joint PDF is low. Thus in this case most of the targets can be distinguished from the background as statistical outliers, but they are *not* outliers with respect to wavelet amplitude alone. Some of the target crosses are located within regions of high background density, but these all have high local energy and thus correspond to targets embedded in strong clutter, in this case IR cirrus, where the targets are very difficult to identify by visual inspection. When in regions of low local energy,

the targets are bright relative to the *local* background and are thus perceived more easily.

The modelling of the joint distribution of y and E is done along similar lines to the joint distribution of T and s in Section 3, to avoid the disadvantages of estimating the joint PDF of more than one random variable mentioned in Section 4.1. Once the self-similarity parameter h has been estimated, y is independent of scale, so the dependence on s can be ignored. Details of the modelling of T and E are given in [14], an outline of which is given here. The idea is to predict the conditional cumulative probability distribution of T given E . Observation of real and simulated intermittent data (e.g. Fig. 3 for the IR celestial background) has shown that this conditional distribution is self-similar, being only dependent on an energy-dependent dilation factor, which can be estimated by calculating the conditional standard deviation of T as a function of E . It has been found that a power-law fits this function quite well, so the following model of conditional threshold exceedance probability has been used:

$$\Pr(y' > y | E' = E) = \exp\left(-\left(\frac{y}{\beta E^r}\right)^c\right) \quad (15)$$

The parameter r is another measure of intermittency of the background, and can be estimated by linear least squares regression of the logarithm of conditional standard deviation versus the logarithm of local energy. The following energy-normalised wavelet amplitude

$$I = TE^{-r} \quad (16)$$

is an energy-independent measure of rarity, and targets and clutter can be discriminated more effectively based on I instead of the un-normalised y . Fig 5 shows the joint statistics of I and E , where it can be seen that the dependence on E is greatly diminished compared with T (Fig. 4), and that a single threshold on I is sufficient to detect most targets with very few false alarms.

The last step is to combine the rarity measures $I(L)$ over multiple regions sizes L . The idea is to select the "best" region size L given the local distribution of clutter. If the object in question is completely embedded in relatively strong clutter then detection will be difficult for any size L . If the object is close to but not embedded in strong clutter then a small L will give the best detection sensitivity, but a larger value is better otherwise. The "best" size corresponds to the largest region not overlapping strong clutter, in which local statistics are most accurately represented.

An overall measure of rarity is therefore defined as follows. First $I(L)$ is normalised by dividing by its standard deviation over all wavelets. Then for each

feature the largest region size L is selected for which the function $I'(L)$ given by:

$$I'(L_n) = I(L_n) - I(L_{n-1}), \quad n \geq 2 \quad (17)$$

is less than 3.5 for all lower scales:

$$L_{\max} = \max\{L_n : I'(L_k) < 3.5 \forall k : 2 \leq k \leq n\} \quad (18)$$

Sudden decreases in $I(L)$ are considered because these correspond to a step increase in local energy, which occurs when a region of clutter is encountered, as required. Decreases of less than 3.5 standard deviations are ignored on the grounds that they are not strong enough to correspond to large changes in background statistics. An overall measure of unusualness is then defined as the mean value of I of size not exceeding L_{\max} :

$$I_{\text{opt}} = \frac{\sum_{L \leq L_{\max}} I(L)}{\sum_{L \leq L_{\max}} 1} \quad (19)$$

Maximum and minimum values of $I(L)$ were also investigated but were found to be less robust discriminators between targets and the background.

Figs. 6a and 6b compare the target detection performance of the original scale-normalised wavelet amplitude T and the new energy-normalised measure I , where in both cases the threshold has been set at 2 standard deviations (of T or I). The use of T results in 21 targets being detected along with 19 false alarms, compared with 29 targets and only 1 false alarm when I is used. The improvement in performance is due to the ability to take local background conditions into account in the neighbourhood of each target.

4.3 Image Sequence Analysis: Track Before Detect

The final example of the use of additional random variables is to aid the detection of targets in time-sequenced imagery, where target motion and the evolution of clutter in time need to be taken into account. Many target detection methods, including most earlier ones [15], are based on "track after detect", in which initial processing is limited to single frames, and track association algorithms are employed subsequently to characterise movement. These methods are not effective when there is insufficient information in single frames to discriminate targets from a large number of clutter objects, typically when the target is amongst clutter of equal or greater edge contrast, or of similar texture, such as roads, buildings, cloud edges and sea glint. Human vision has the same limitation; objects which cannot be recognised in still frames can often be detected subliminally in a movie by virtue of movement or other distinguishing temporal behaviour.

We consider instead a "track before detect" (TBD) method, which combines image data from different frames prior to detection, treating the data as a single 3-dimensional set, and applying 3D filters. In this context, movement is inherent in the 3D orientation of geometric features within the data, and directly affects the response of 3D filters. MFT is applicable to the analysis of 3D image sequence data [16,17], using the same underlying principles as for 1D signals and 2D images. The PSD is a 3D function, providing information about both spatial and temporal correlation of the background.

In the case of point targets against a bland background, where uncorrelated sensor noise dominates, MFT results in a classic velocity-tuned filter which integrates image intensity along the trajectory of the target. A very different filter is required for scenes with strong static clutter but little sensor noise. The background is usually correlated in space (e.g. with a $1/f$ PSD) and very strongly correlated in time (nearly constant), so MFT leads to mean subtraction over time and a decorrelation filter in space, e.g. a (fractional) Laplacian filter [16]. Additional spatial smoothing is required for larger targets. For both types of background, however, the matched filter is strongly dependent on velocity for small targets, so often a large number of filters are required for low targets with low SNR.

We have been able to take advantages of symmetry in static downward-looking imagery to simplify the background statistical models, using invariance with respect to translation, rotation and scale. Unfortunately, the statistics of space-time structure in image sequences are often much more complex, with less natural symmetry. Whereas the two spatial dimensions often have an equal footing and similar statistics, the time dimension is generally different, for example a scene may be very highly correlated in time but much less so in space. The space-time equivalent of rotation is (approximately) a Galileian transformation, i.e. a velocity shift. The statistics of space-time structure are usually strongly dependent on velocity shifts, unlike spatial rotation. The effect of time dilation is usually quite different from spatial dilation, and both usually affects the statistics of movement dramatically, as can be seen when a movie is played at the wrong speed. In forward-looking imagery the situation is even worse, as there is no longer invariance with respect to the vertical image co-ordinate, or to spatial orientation.

The complexity of space-time image structure can be seen with forward-looking imagery of sea glint (Fig. 7). A bank of simple $3 \times 3 \times 3$ space-time bar filters has been applied to pick out bars with 13 different orientations. Their statistics are very complex for sea glint, as shown by Fig. 8, where the thresholds for an exceedance probability of 10% are compared. These filters have been normalised to give identical exceedance rates in Gaussian white noise. With this forward-looking imagery

much of the spatial symmetry is lost also, as the clutter strength is strongly dependent on the vertical image coordinate y and also on spatial orientation, with stronger returns from horizontal bars and edges than vertical ones.

The wavelet transform has fewer advantages when applied to time-sequence imagery than to static images because of the above lack of symmetry. The conventional 3D wavelet transform applies a single dilation operator, whereas space and time should be scaled independently. It is possible, however, to employ a non-homogenous form of 3D wavelet analysis [18] to scale space and time independently.

The approach of using a bank of matched filters involves looking for unusually strong responses from each filter, regardless of the responses from the other filters, and so requires knowledge only of the univariate statistics of single filter responses. Instead, we consider the multivariate statistics of a small number of filters which individually are not as good discriminators of targets and clutter, but when considered jointly can be more effective at this task. As with static imagery, we consider a small number of simple geometric features, oriented bars in space-time, which in this case represent moving blobs, where 3D orientation corresponds to the velocity of motion. The aspect ratio of these bars (ratio of length to width) is small, about 3, so the resolution in velocity is low.

We now consider what additional filtering will aid clutter rejection, in the same way that local energy was used for intermittent static backgrounds (Section 4.2). Conventional target detection methods [19] perform poorly on sea glint compared to human vision, because the target detection filters also respond well to sea glint, not recognising its transience. It is easier to recognise the target because it *doesn't* respond well to a transient detection filter, than because it triggers a filter tuned to the target. This leads us to consider the joint statistics of 3D bar filters, weakly tuned to velocity, and simple non-linear filters measuring the variability in brightness along the length of each bar:

$$V = \max_k \left(\left| x_k - \text{med}_k(x_k) \right| \right) \quad (20)$$

where x_k are the pixel values along the central axis of the 3D bar. This filter will give a strong response to intermittent sea glint (flashes) but usually a much weaker response to a moving target.

The joint statistics of a pair of these filters are shown in Fig. 9a for sea glint, where an incoming point target has been simulated which is significantly fainter than the glint. The target has been marked in Fig. 9a with a cross, and lies well outside the background distribution, but like local energy (Section 4.2) it is only by considering the joint statistics of these filters that discrimination

between the target and clutter is possible. In some situations only targets coming head-on, with no apparent motion, will be of interest. Where lateral motion is also expected, the above joint statistics could be studied independently for filter pairs, each tuned to a different velocity. However, with the extra discrimination provided by the variability filter, only a low resolution in velocity is required, and hence a small number of filter pairs.

Additional discrimination is available by adding a third measurement to the joint statistics, the vertical image coordinate Y , to take account of the reduction in sea glint strength nearer the horizon, and also the very different space-time structure above the horizon (Fig. 9b). The joint statistics of the bar filter, variability and Y are not well suited to a parametric model like that involving local energy (Section 4.2), partly because there are arbitrary discontinuities, for example at the horizon. Instead a non-parametric method of conditional threshold exceedance probability estimation is used, described briefly next. Details are available in [19].

The idea is to estimate the conditional threshold exceedance probability of a single target-related measurement T given a set of background-related measurements B . In our case the target measurement is the bar filter output, and the background measurements are V and Y . This method is also applicable to Section 4.2, where the background measurement is local energy E . The conditional probability can be written:

$$\Pr(T > T_0 \mid B = B_0) \quad (21)$$

We consider conditional rather than joint probability because the marginal probability density of the background measurements are assumed not to be of interest; for example if V is unusually large, this does not imply that a target is very likely, only that a transient glint of unusual strength occurs. If, on the other hand, the bar filter is unusually strong *given* a measured level of glint, a target is more likely. The values of B are of no significance, other than in conditioning the value of T .

Estimating the probability given by equation (21) requires some sort of local averaging in B , using a window, assuming the latter has a continuous distribution. The following is a simple estimator of this probability:

$$\frac{\sum_{T_i > T_0} \exp(-A \|B_k - B_0\|)}{\sum_k \exp(-A \|B_k - B_0\|)} \quad (22)$$

where the exponential function is a Parzen window which gives greater weighting to points which are close to B_0 than those further away, and A is the reciprocal of the radius of the window. As with probability density

estimation, when selecting A there is a trade-off between high resolution (A is small) and statistical accuracy (A is large). We adapt A according to the density of values B in the neighbourhood of B_0 , so that a large radius is used in sparse regions and a small radius in dense regions. This is achieved by choosing A such that the denominator of equation (22) is a prescribed value which can be thought of as the number of measurements being taken into account when estimating the conditional probability.

Equation (22) is an unbiased estimator of equation (21) only if the true density is independent of B within the region of influence of the Parzen window. Where there is global independence, there is no need for a window function; estimates of unconditional exceedance probability will suffice. If there is dependence on B , then at most points the gradient of exceedance probability will be non-zero, and thus linear dependence will be a better model than a constant value (local independence). This can be seen clearly in the example of Fig. 9a, where contours of constant density are not parallel to the vertical axis, but appear to be smooth, implying a linear model would be a major improvement over a constant model.

A linear model is defined using the following formula:

$$\frac{\sum_k \chi(T - T_0 + L(B_k - B_0)) \exp(-A \|B_k - B_0\|)}{\sum_k \exp(-A \|B_k - B_0\|)} \quad (23)$$

where L is a linear function of the random variables in B and χ is the function which returns 0 if the input is less than 0 and 1 otherwise. The coefficients of the function L are estimated from the data using linear least squares regression but where the data values are weighted according to the similarity of B to B_0 and the magnitude of T , where there is a bias towards large values of T , so that attention is given to the tails of the distribution. Once L is estimated, equation (23) is used to estimate the conditional threshold exceedance probability.

Figs. 10a and 10b show the 10 events with the lowest threshold exceedance probability, estimated using the above method, both on the scatter plot of T against V , and their corresponding locations within the image. The sea target is the 2nd most unusual event, with an extremely low estimated exceedance probability of 1.7×10^{-29} . Without the conditioning with respect to the background variables, the target exceedance probability would have been much higher, with the sea glint dominating the results. From a visual inspection of the scatter plots (Figs 9a and 9b), the target appears to be the least probable event, but the point with greatest rod filter output has been assigned an exceedance probability of zero. This is a technical failing of the estimation algorithm; a hard decision boundary is used in the

numerator of equation (23), so whilst uncertainty in the background random variables Y and V is taken into account in the Parzen window, there is as yet no analogous uncertainty in the target variable T . If there happen to be no sample values to the right of the linear decision boundary, the current estimate of exceedance probability is therefore zero. This limitation could be rectified by using a soft decision boundary, for example using the logistic function instead of χ .

As with many detection methods, temporal consistency can be used as an additional criterion to improve performance further. Although in this example the target has been detected successfully from just a single snapshot of 5 frames, if the SCR was much less, the exceedance probabilities of the most intense sea glint could be as low as that of the target. When the above analysis is repeated on later groups of frames, the target continues to have a very low exceedance probability, whereas the locations of the other low-probability events within the sea change. Fig. 11 shows the top 40 detections superimposed from each of four successive time intervals, in which it can be seen that the only persistent events are the target, a few points on the land (which are unchanging), plus 3 points within the sky (which are known to be dead pixels and therefore constant). Out of these persistent events, the estimated target probability (ranging between 10^{-8} and 10^{-29}) is much lower than that of the other events (typically 10^{-3} to 10^{-5}).

7. References

1. V. Cappellini, A.G. Constantinides, P. Emiliani, *Digital Filters and their Applications*, Academic Press, 1978.
2. W.B. Davenport, W.L. Root, *An Introduction to the Theory of Random Signals and Noise* McGraw-Hill, 1958.
3. B.B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, 1982.
4. J. Feder, *Fractals*, Plenum Press, 1988.
5. R.F. Voss, "Fractals in Nature: From Characterisation to Simulation", Chap. 2 in *The Science of Fractal Images*, H.O. Peitgen and D. Saupe, eds., Springer-Verlag, New York, 1988.
6. V.S. Vladimirov, *Generalised Functions in Mathematical Physics*, Mir Publishers, Moscow, 1979.
7. W.H. Press, *Numerical Recipes in C: the Art of Scientific Computing*, Cambridge University Press, 1992.
8. J.G. Jones, R.W. Thomas, P.G. Earwicker, "Multiresolution Analysis of Remotely Sensed

- Imagery", *International Journal of Remote Sensing*, 12(1), pp.107-124, 1991.
9. J.G. Jones, G.W. Foster, A. Haynes, "Fractal Properties of Inertial-Range Turbulence with Implications for Aircraft Response", *Aeronautical Journal*, pp.301-308, Oct 1988.
 10. J.G. Jones, R.W. Thomas, P.G. Earwicker, S. Addison, "Multiresolution Statistical Analysis of Computer-Generated Fractal Imagery", *CVGIP: Graphical Models and Image Processing*, 53(4), pp. 349-363, 1991.
 11. A.N. Kolmogorov, "A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible flow at high Reynolds number", *J.Fluid Mech.*, 13(82), 1962.
 12. V.N.Vapnik, *Statistical Learning Theory*, Wiley, 1998.
 13. B.S.Everitt, *Cluster Analysis*, Wiley, New York, 1974.
 14. G.H.Watson and S.K.Watson, "Detection of unusual events in intermittent non-Gaussian images using multiresolution background models", *Optical Engineering*, 35(11) pp.3159-3171, Nov 1996.
 15. B.Bhanu, "Automatic Target Recognition: State of the art survey", *IEEE Trans. on Aerospace and Electronic Systems*, 22(4), pp.364-379, July 1986.
 16. I.S.Reed, R.M.Gagliardi and H.M.Shao, "Application of Three-Dimensional Filtering to Moving Target Detection", *IEEE Trans. on Aerospace and Electronic Systems*, 19(6), pp.898-905, Nov 1983.
 17. I.S.Reed, R.M.Gagliardi and L.Stotts, "Optical Moving Target Detection with 3-D Matched Filtering", *IEEE Trans. on Aerospace and Electronic Systems*, 24(4), pp.327-335, July 1988.
 18. T.J.Burns, S.K.Rogers, M.E.Oxley, D.W.Ruck, "A Wavelet Multiresolution Analysis for Spatio-Temporal Signals", *IEEE Trans. on Aerospace and Electronic Systems*, 32(2), pp. 628-649, April 1996.
 19. G.H.Watson and S.K.Watson, "Detection and Clutter Rejection in Image Sequences based on Multivariate Conditional Probability", *SPIE Proceedings*, vol 3809, 1999.

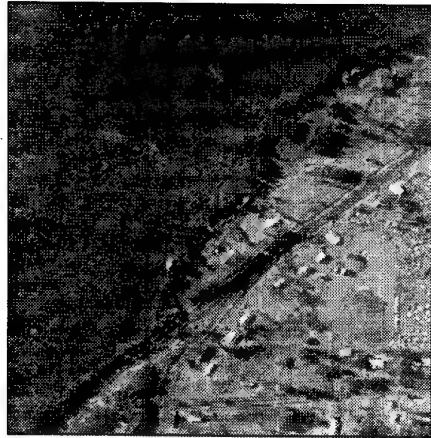


Fig 1a: TIMS image with roads and buildings

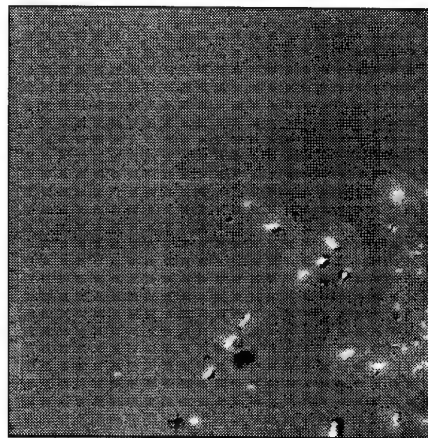
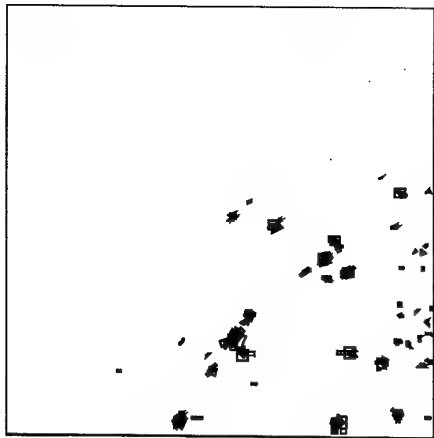
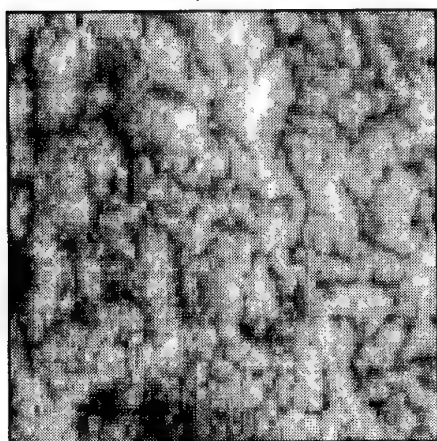


Fig 1b: Feature characterisation and reconstruction at high amplitude threshold

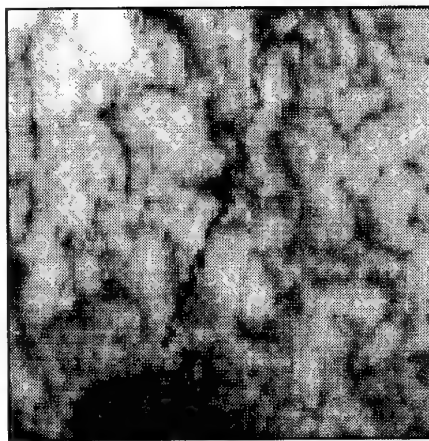


Fig 1c: Feature characterisation and reconstruction at low amplitude threshold

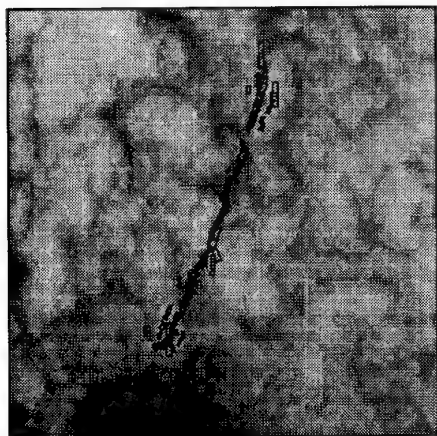
Fig. 2: Detection of ship track in AVHRR imagery using vector wavelets



(a) Band 1 (visible)



(b) Band 3 (near IR), with ship track



(c) Detections using wavelet-based
Mahalanobis distance



(d) Detections using pixel-based
Mahalanobis distance

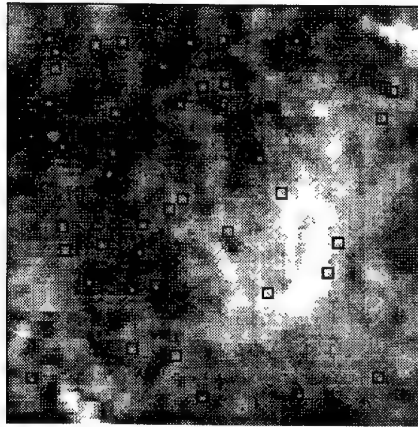


Fig. 3: IRAS image of celestial background imagery with simulated point targets embedded

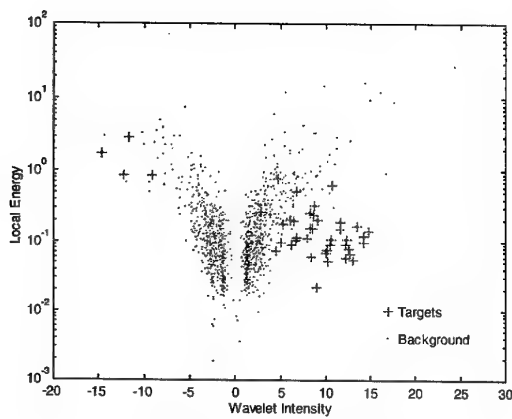


Fig. 4: Joint distribution of wavelet intensity (T) and local energy (E)

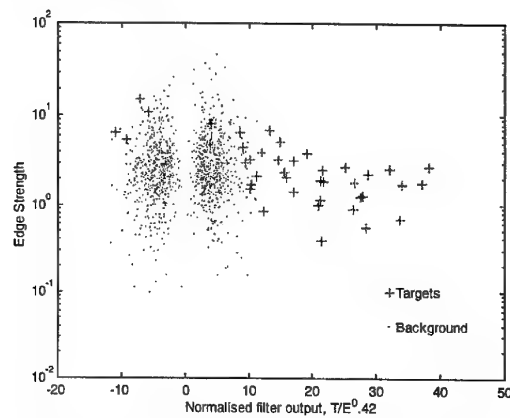


Fig. 5: Joint distribution of energy-normalised wavelet intensity ($T/E^{0.42}$) and local energy (E)

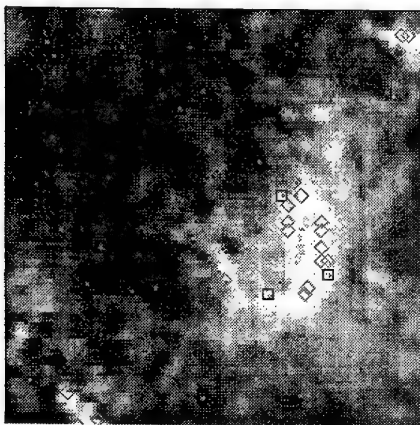


Fig. 6a: Detections at a threshold of 2σ based on wavelet intensity (T)

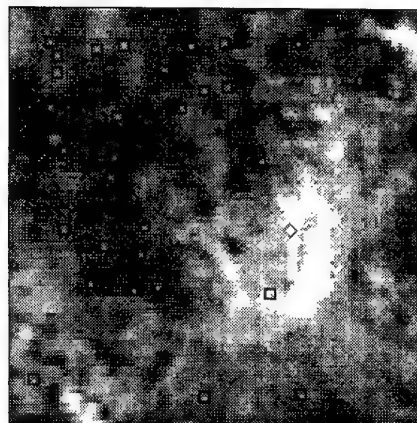


Fig. 6b: Detections at a threshold of 2σ based on energy-normalised wavelet intensity (I)

Squares = embedded targets, diamonds = false alarms



Fig. 7: Sea glint data
Single-pixel slowly-moving target (arrowed) artificially embedded in sea clutter

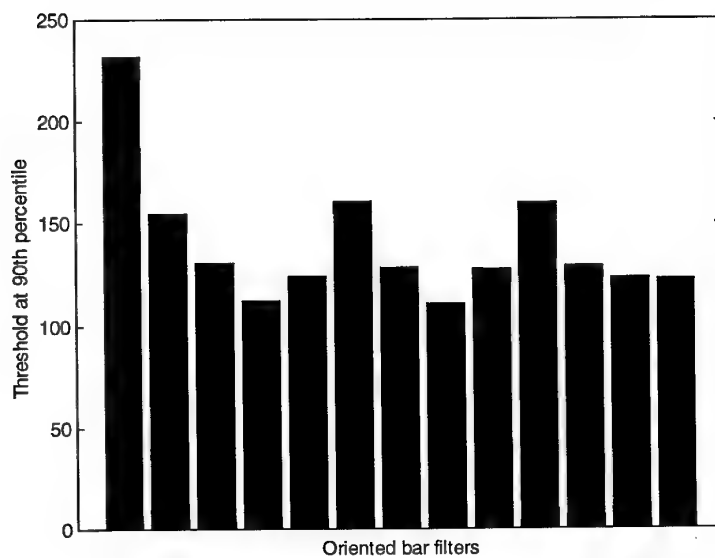


Fig. 8: Statistics of responses of simple 3x3x3 space-time bar filters

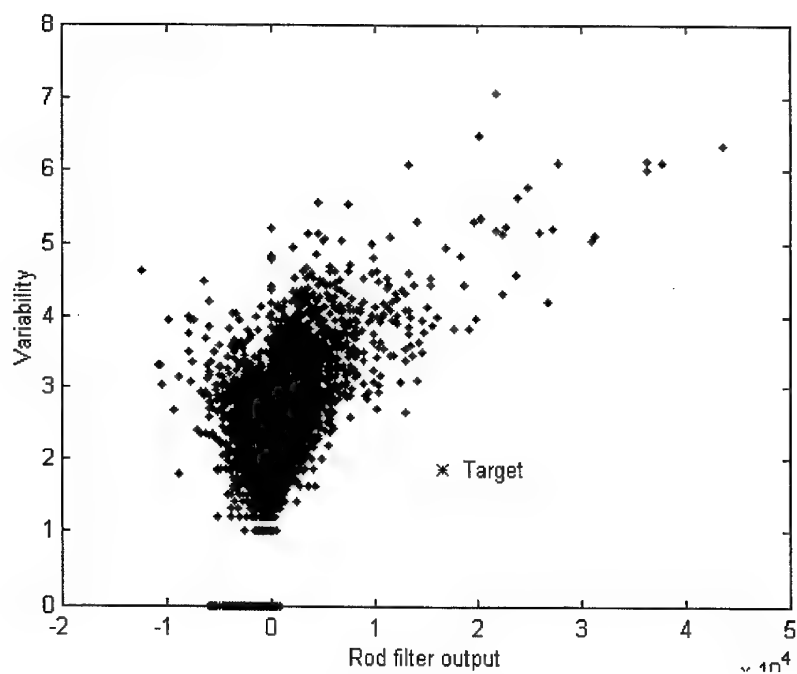


Fig. 9a: Joint statistics of rod filter output (T) and variability measure (V)

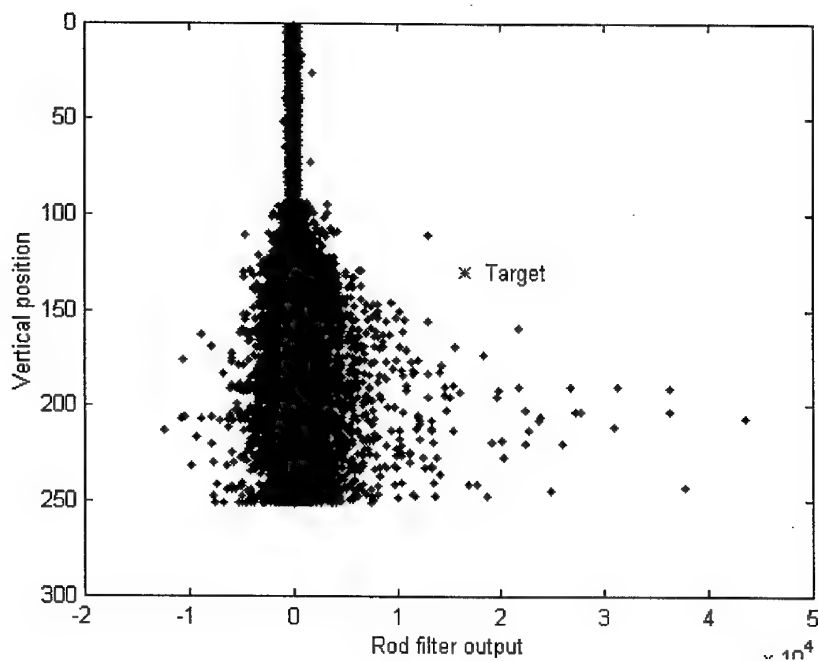


Fig. 9b: Joint statistics of rod filter output (T) and vertical position (Y)

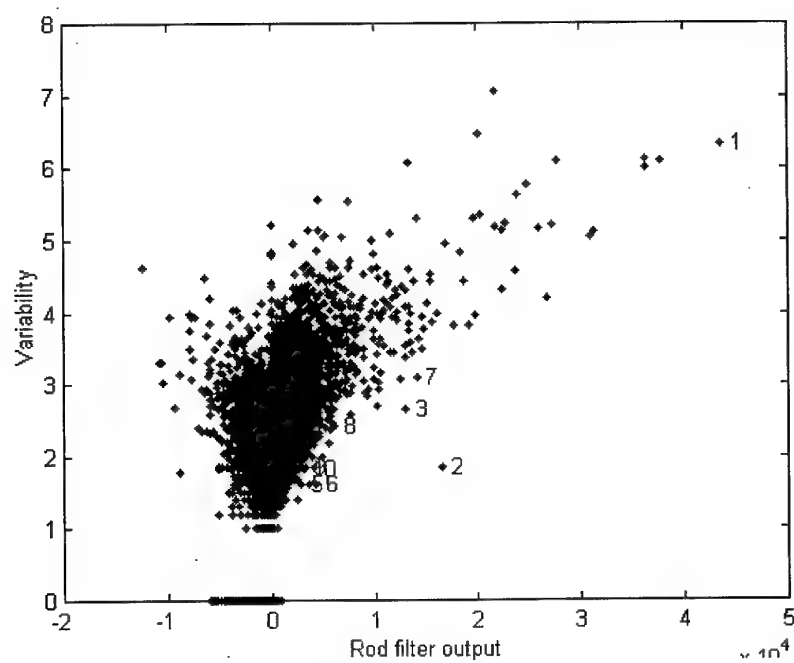


Fig 10a: The 10 most unusual events as a result of 3-D statistical analysis, shown within the joint statistics of rod filter output (T) and variability measure (V)

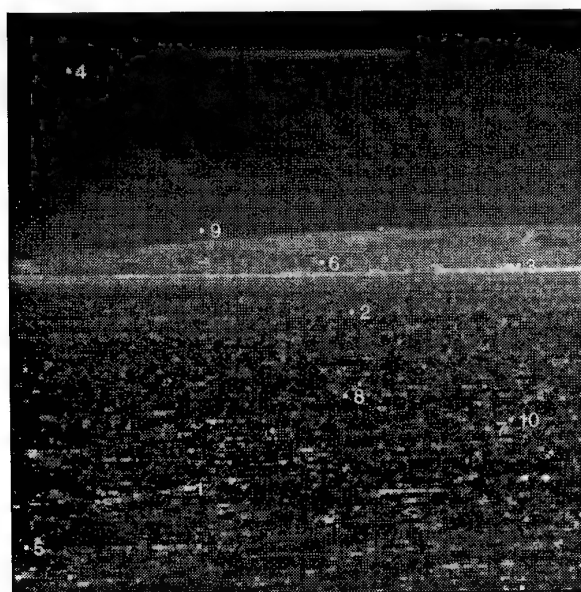


Fig 10b: Locations of the 10 most unusual events from 3-D statistical analysis

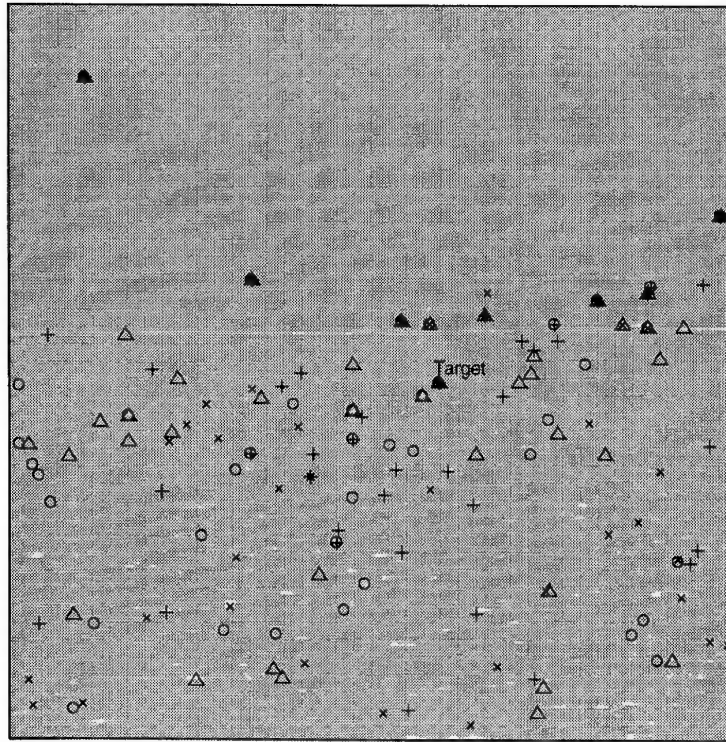


Fig. 11: Demonstration of lack of persistency of sea glint
Top 40 detections within 4 successive time intervals superimposed
Detections marked with plus, cross, circle and triangle respectively

Invariant Feature Extraction in Wavelet Spaces

G.H.Watson

Room 1052, A2 Building, DERA Farnborough, Ively Road, Farnborough, Hants, GU14 0LX, UK

1. Introduction

This paper is concerned with the extraction of geometric information from signals and images which is suitable for pattern recognition, but which is as insensitive as possible to the conditions of data collection, in particular sensor noise, background clutter and changes in the geometric relationship between the sensor and the scene (e.g. 3D viewing aspect). We describe a method of feature extraction which involves extending wavelet analysis beyond translation and dilation, leading to invariance to a greater variety of transformations, as well as insensitivity to noise and clutter. Feature extraction involves searching for local maxima in a generalisation of the continuous wavelet transform, and we describe how Riemannian geometry is used to aid this process. Additional topics covered are signal and image reconstruction, redundancy removal and application to 3D aspect-invariant target recognition.

The analysis of any data, including signals and images, involves the extraction of application-specific information and the rejection of other less relevant data. Transforming the data into a form where the information of interest is easier to obtain often facilitates the analysis. Clearly, general-purpose transformations are more useful if they transform the data into a form of information that is relevant to many applications. This is the case with the Fourier transform because many phenomena of interest to scientists and engineers are more simply described in terms of sinusoids than the raw data samples. The wavelet transform [1,2] is another example, which has become increasingly popular in recent years because of its ability to identify both position and scale, and hence the role it plays in multi-resolution signal and image processing.

The motivation behind the material covered in this paper is the desire to recognise objects of interest, but to ignore other information not of interest, and to segregate the two as efficiently as possible. However, as with the Fourier and wavelet transforms, we wish the method of analysis to be as generic as possible, so rather than attempting a high-level characterisation of the sensor data, for example the classification of military targets, a lower-level characterisation in terms of simple geometric features such as sine waves or compact wavelets is sought. The aim is to represent application-specific data as combinations of these lower-level features.

There are two categories of data which we wish to ignore, or at least separate from the object of interest:

- Noise and clutter. We aim to ignore this by selecting a subset of the above geometric features, related to the information of interest, and rejecting the remainder. The feature decomposition should be designed to facilitate this selection.
- Information describing the geometric relationship between the sensor and the information of interest, not the information itself. For imaging sensors this comprises range, bearing and 3D orientation of the objects being viewed. For acoustic and radar sensors this comprises range, time of arrival, multipath, Doppler shift and possibly other effects such as distortion. These variables are sometimes called nuisance parameters. We aim to reject this information by constructing canonical forms (Section 6) which do not depend on this geometrical relationship.

We thus conceptually segregate the signal or image data into information of interest, the foreground, and the remainder, the background. The extraction of information from the data is now analogous to the detection and characterisation of the foreground and the rejection of the background, where a detection occurs when a foreground feature is extracted from the data. The extracted features can therefore be interpreted as the output from a set, or *dictionary*, of detection filters applied to the signal or image. This dictionary depends on the application; the foreground and background should be easy to segregate based on the outputs of the filters it comprises. The Fourier dictionary, sines and cosines (or complex exponentials), is suitable for foregrounds with strong harmonics, and has the additional advantage that the amplitudes of the Fourier coefficients are translation-invariant. The wavelet dictionary comprises all translations and dilations of the mother wavelet, and so is suitable if the position and scale of the foreground objects in the signal or image are not known a priori.

In general, an *economical* but *complete* representation of the signal or image is sought in order that the information of interest is easily extracted but still fully represented. A conventional approach to achieving economy and completeness is to use a fixed basis for the

dictionary so that the component functions span the signal space and are linearly independent, as with the discrete Fourier and wavelet transforms. However, the use of a fixed basis constrains the size of the dictionary, limiting the scope and sensitivity of the signal or image representation.

More recent developments have therefore considered adapting the dictionary functions separately to each signal or image rather than using a fixed dictionary for all signals. In Coifman and Wickerhauser [3], the master dictionary is a specially constructed library of orthonormal bases. An orthonormal dictionary is adapted from this library in order to minimise the information content in the representation of the signal. More generally [4,5], the master dictionary is non-orthogonal, but a smaller subset is similarly adapted to the signal to maximise the efficiency of the representation. This approach is aimed at resolving the conflict between flexibility, where a large dictionary would ordinarily be required, and economy, and has been adopted both in the Wavelet Packet Transform of Coifman et al [4] and also in the Matching Pursuit algorithm of Mallat et al [5], where in both cases the functions are chosen or generated to give the best approximation to the signal or structures within the signal.

This paper considers a variant of the Matching Pursuit algorithm, in which the master dictionary is a manifold of infinitely many functions. Like Matching Pursuit, the idea is to generalise the wavelet transform, so that the dictionary is generated from a mother wavelet, but includes transformations other than translation and dilation. This enables the efficient and invariant representation of a greater variety of foreground information, for example with variable orientations, affine transformations, frequency shifts and chirp angles. Because the dictionary is generated by the action of a continuous group of transformations on a mother wavelet, for convenience the method is denoted shape adaptive wavelet analysis, the dictionary is called a wavelet space, the constituent functions are called wavelets, and correlation of the signal or image with these functions is called the shape adaptive wavelet transform (SAWT).

We adapt the wavelets to the signal or image data by select only those whose SAWT are local maxima, in a manner analogous to Mallat and Hwang [6]. We show that these functions have three useful properties:

- They are the best local approximations to the signal or image [6], thus enhancing the quality of the representation and reconstruction of the original data (Section 2.4).
- They represent structures with locally maximal signal-to-noise ratio (SNR), and are thus well suited to represent foreground information (Sections 2.3 and 2.4).

- The features extracted from local maxima are invariant under the transformations associated with the wavelet space, for example translation, rotation and dilation. This property aids sensor-invariant pattern recognition (Section 6).

The selected wavelets are not in general orthogonal, but a method of redundancy removal is presented (Section 5) which increases parsimony and also simplifies the process of reconstructing the foreground.

The remainder of this paper is laid out as follows. Section 2 describes the underlying principles of the wavelet space, the extraction of geometric features at local maxima in the SAWT, and how these are related to matched filters. Section 3 gives an overview of the methods covered in this paper, including feature extraction, signal or image reconstruction and sensor-invariant representation. Section 4 gives details of feature extraction, including the use of Riemannian geometry and local co-ordinate transformations to improve efficiency, and the use of shape-adaptive wavelets for characterising foreground objects. Section 5 discusses redundancy removal and signal or image reconstruction. Section 6 describes how canonical forms are constructed from the above feature decompositions, which are independent of viewing aspect and related sensor settings, and their use in aspect-invariant recognition of aircraft viewed from an imaging sensor.

2. Shape Adaptive Wavelet Analysis

2.1. Matched filters and dictionaries

In this paper we assume all signals and images are real or complex valued functions defined on the real line (e.g. acoustic signals), a 2D plane (e.g. grey-level images), or a higher-dimensional space (e.g. image sequences). We will develop the theory for complex functions, but the theory of real-valued functions is very similar. Expressed formally, we consider functions which are members of the Hilbert space F_n of Lebesgue square-integrable functions $\{f: \mathbf{R}^n \rightarrow \mathbf{C}\}$ with the inner product

$$\langle f_1, f_2 \rangle = \int_{\mathbf{R}^n} f_1(u) \overline{f_2(u)} du \quad (1)$$

for any two functions $f_1, f_2 \in F_n$ and $u \in \mathbf{R}^n$. We are interested in a dictionary of filters which respond well to foreground information. Here we consider linear filters only, each of which can be written as the following inner product:

$$\langle f, w \rangle = \int_{\mathbf{R}^n} f(u) \overline{w(u)} du \quad (2)$$

where $f \in F_n$ is the signal or image, $w \in F_n$ is the filter, n is the dimension of the signal or image and $\langle f, w \rangle$ is the output of the filter. We use matched filter theory [7] to derive linear filters w which will optimally detect a given foreground signal f when embedded in background noise with power spectral density N :

$$W(\omega) = \frac{F(\omega)}{N(\omega)} \quad (3)$$

where F and W are the Fourier Transforms of f and w respectively. The matched filter w is, however, as specific in position, scale and shape as f . To detect (i.e. distinguish from the background) a wide variety of foreground structures requires a large dictionary of filters $\{w\}$, so in this paper we consider dictionaries with a manifold structure, where the dimension is typically 3 or more.

2.2. The wavelet space and SAWT

Shape adaptive wavelet analysis transforms a signal or image into a linear combination of dictionary functions, in common with the Fourier, Wavelet and Wavelet Packet Transforms. The dictionary, called the *wavelet space*, is a C^2 manifold (continuously differentiable to second order) of infinitely many functions called *wavelets* which are chosen to enable us to detect structures in the signal or image which provide information of interest. Because the wavelet space is a manifold, for each wavelet w in the wavelet space there is a chart (co-ordinate system) $c: W \rightarrow R^n$ which maps the wavelet to a vector $x = c(w)$ where the elements in this vector are the wavelet's co-ordinates with respect to the chart c . We have a lot of freedom of choice of the chart, and as we shall see in Section 4, carefully constructed charts benefit feature extraction.

We use the inner product given in equation (2) to define the following real- or complex-valued function on the wavelet space,

$$T_f(w) = \langle f, w \rangle, \quad (4)$$

which in the context of this paper is called the *Shape Adaptive Wavelet Transform* (SAWT). This function, analogous to the continuous wavelet transform, represents, at least in principle, the action of the entire dictionary of filters in the wavelet space W on the signal or image f .

Examples of transforms and 'wavelets' in common use are:

- The continuous Fourier Transform [8], where the 'wavelets' $\{w\}$ are (untruncated) sinusoids parameterised by frequency.

- The continuous Wavelet Transform [1,2], where the wavelets $\{w\}$ are translated and scaled versions of the mother wavelet, parameterised by position and scale:

$$w_{p,s}(t) = s^{-\frac{1}{2}} w\left(\frac{t-p}{s}\right) \quad (5)$$

- The Short Term Fourier (Gabor) Transform [1,2], where the wavelets are truncated sinusoids, typically multiplied by a windowing function h , and are parameterised by position and frequency:

$$w_{p,f}(t) = h(t-p) e^{2\pi i f(t-p)} \quad (6)$$

- The Hough Transform [9] for images, where the wavelets are traditionally taken to be lines parameterised by gradient and intercept.

2.3. Whitening

Each of the above transforms can be used to optimally detect signal structures which approximate the respective dictionary functions when embedded in uncorrelated (white) noise, because for this type of background the matched filter is identical to the signal or image structure being sought. Thus the Fourier Transform is an optimal detector of sinusoids, the Wavelet Transform is an optimal multi-resolution detector for the mother wavelet, the Gabor Transform is an optimal detector for truncated sinusoids and the Hough Transform is an optimal detector for linear features in images. Where the background is correlated (not white) the wavelet filters $\{w\}$ can be interpreted as the optimal detectors (matched filters) for the related signal or image components $\{v\}$ given by:

$$V(\omega) = W(\omega)N(\omega) \quad (7)$$

where W and V are the Fourier Transforms of w and v respectively, and N is the power spectral density of the noise. In practice it is much easier to apply an invertible spectral transformation to the signal or image to whiten the noise before applying the shape adaptive wavelet transform. In the Fourier domain, such spectral transformations have the form

$$F'(\omega) = \frac{F(\omega)}{M(\omega)} \quad (8)$$

where F is the unwhitened signal or image, F' is the whitened version of F and where M is chosen to satisfy $|M(\omega)|^2 = N(\omega)$ with N the background noise. Henceforth, we therefore assume that the background is pre-whitened.

2.4. SAWT, SNR and approximation error

We now explain how the wavelet space and the SAWT are used to extract foreground information from a given signal or image. Suppose our given signal or image is a Gaussian white noise random process. For a given point (wavelet) w in the wavelet space, the variation of the SAWT at that point will also be Gaussian with zero mean, so the statistics of the response are characterised entirely by standard deviation. This in turn is proportional to the L_2 norm of w :

$$\sigma(N) \propto \sqrt{\int_{\mathbb{R}^n} |w(u)|^2 du} = \|w\|_2 \quad (9)$$

If all the wavelets in the wavelet space have the same L_2 norm, then those wavelets $\{w_k\}$ at the local maxima in the modulus of the SAWT are seen to represent the signal components with the locally highest SNR, which can be interpreted as being most likely to belong to the 'foreground' of interest and not to the 'background' if the statistical distribution of the SNR values is being considered. We therefore normalise all the wavelets in the wavelet spaces considered in this paper to unit L_2 norm. This is why there is a scale factor s dividing the conventional position-scale wavelet transform.

Not only do the local maxima represent signal components most likely to be foreground, but, as we show next, they are also the best local approximations to the signal data. In considering such approximations we need to define an appropriate measure of approximation error. We require the approximation of the signal represented by the wavelet expansion to be stable with respect to the background. Differences between signals should therefore be measured in terms of their likelihood of arising at random from the background, so that "small" errors can be disregarded as being of little significance. In the case of Gaussian white noise the L_2 norm of the difference between two signals is such a measure of error, and the one that we will adopt in the remainder of this paper. For correlated backgrounds the approximation error is derived by whitening the difference between the signals prior to the calculation of the L_2 norm. This error is also the maximum signal-to-noise ratio of the residual when regarded as a signal embedded in unit energy noise. Suppose any filter w is employed to detect the residual f . Then the signal to noise ratio is given by:

$$\text{SNR} = \frac{|\langle f, w \rangle|}{\|w\|} \leq \|f\| \quad (10)$$

where we have used the Cauchy-Schwarz inequality. Thus the L_2 norm is a measure of the difficulty of detection of the residual using a matched linear filter. If this quantity is small, it is natural to regard the residual as being small, and the signals as being similar.

2.5. Approximation with the SAWT

We now show that the best local approximation to the signal or image is given by the wavelet at a local maximum in the SAWT multiplied by the value of the SAWT. First we consider the best approximation of the signal f with a wavelet w multiplied by the complex amplitude A , where the approximation error is given by:

$$E_f(w, A) = \int_{\mathbb{R}^n} |f(u) - Aw(u)|^2 du \quad (11)$$

We temporarily fix w and consider the minimum of $E_f(w, A)$ with respect to A .

Expanding equation (11) we obtain:

$$E_f(w, A) = \|f\|_2^2 + |A|^2 \|w\|_2^2 - 2\text{Re}(\overline{A}T_f(w)) \quad (12)$$

We recall $\|w\|_2 = 1$, because the wavelets are normalised to unit L_2 norm, and we write $A = ae^{i\theta}$ and $T_f(w) = be^{i\phi}$. Equation (12) then reduces to

$$E_f(w, A) = \|f\|_2^2 + |A|^2 - 2ab \cos(\theta - \phi) \quad (13)$$

We require a minimum of $E_f(w, A)$ with respect to both a and θ .

For θ , we see by inspection that $\theta = \phi(\pm 2n\pi)$ and (13) then becomes

$$E_f(w, A) = \|f\|_2^2 + a^2 - 2ab \quad (14)$$

For a , we partially differentiate $E_f(w, A)$ with respect to a and equate to zero to get:

$$a = b; \quad A = T_f(w); \quad E_f(w, A) = \|f\|_2^2 - |T_f(w)|^2 \quad (15)$$

Thus the local minima of $E_f(w, A)$ occur at the same locations in the wavelet space as the local maxima of the modulus of the SAWT, and the optimal amplitude A is then given by the SAWT value at this point.

The equivalence between local maxima of the SAWT and local minima of the error measure is of considerable benefit when searching for local approximations of the signal. Our local approximation to f is given by Aw , so to find the best wavelet Aw directly, we would need to search over the whole space $C \times W$; however, the equivalence of local maxima of the SAWT and local minima of the error measure allows us to search merely for the best wavelet w . The amplitude A is then given by the value of the SAWT at this point. We can thus search over a space with dimension $\dim(W)$ instead of $2 + \dim(W)$, and as search times increase exponentially with dimension, this is of considerable benefit.

3. An overview of the method

The first stage in shape adaptive wavelet analysis is the search for local maxima of the SAWT. This search is conducted as follows:

- Conduct a grid search for local maxima of the SAWT over a low-dimensional sub-manifold of the wavelet space.
- Locally optimise in the full wavelet space each of the local maxima found above.
- Record the amplitudes and positions in the wavelet space of the local maxima as discrete features, for application-specific analysis.
- If background rejection is required, select only the features that belong to the foreground. With L_2 normalisation, the simplest method is to set a threshold on the SAWT; this is best for Gaussian noise. For more intermittent backgrounds the selection criteria are more complex, and covered in [10].
- If sensor-invariant recognition is required, convert the discrete features into a canonical form in which 'nuisance parameters' are removed.

We describe the first two stages in detail in Section 4 and show how consideration of the geometry of the wavelet space enhances the efficiency of the search. The fifth stage is covered in Section 6. If accurate reconstruction of the signal or image data is required, the following additional stages are also involved (Section 5):

- Redundancy removal. The wavelets at the local maxima in the SAWT are not always orthogonal, so in general there is redundancy in the SAWT at the local maxima. Although redundancy removal is required for accurate signal reconstructions, it is often not required in further analysis, for example in detection and sensor-invariant recognition.
- Wavelet subtraction. This method extends the wavelet representation to represent the signal or image to any required accuracy, allowing full inversion of the transform in the limit. This step is usually not necessary when background rejection is involved.

4. Feature extraction

We now describe how the signal or image is decomposed into a discrete set of features, each a function from the wavelet space located at a local maximum in the SAWT. There are two stages:

- Conduct a grid search for local maxima of the SAWT restricted to a low-dimensional sub-manifold V of the wavelet space W . This sub-manifold could be the entire wavelet space if sufficient computation is available, but in higher dimensions this is generally not the case.
- For each local maximum above a given threshold (usually dictated by ambient noise) in the above grid conduct a local search for the nearest local maxima of the SAWT in the full space W .

This method, of course, will not guarantee the detection of all the local maxima of the SAWT, but this guarantee would necessitate limiting the dimension of the wavelet space for reasons of computational feasibility, which is no better. The local search provides the opportunity of improving the approximation to the signal or image without increasing the search time exponentially.

A truly global search over any non-trivial sub-manifold is not possible (unless the SAWT is band-limited) because there are an infinite number of SAWT values to calculate; hence, a finite grid is chosen in such a way that the SAWT at other points can be predicted approximately. To do this efficiently, however, we first endow the wavelet space with a Riemannian metric that quantifies the expected variability of the SAWT. We then construct such a metric and show how it can be used to determine the size of the grid (Section 4.3) and also aid the subsequent local optimisation, considered next.

4.1. Local optimisation

There are numerous "off-the-shelf" local optimisation methods reported in the open literature [11,12] and implemented in computer software, whose effectiveness depend strongly on the type of cost function to be optimised. We take advantage of the second-order differentiability of the wavelet space (Section 2) to use a differential search method which is similar to the method of conjugate gradients [11], but where we take account of the above Riemannian metric. Not only is the SAWT smooth, but its partial derivatives can be computed analytically by differentiation of equation (2):

$$\frac{\partial T_f(w)}{\partial x_i} = \frac{\partial}{\partial x_i} \left(\int_{\mathbf{R}^n} f(u) \overline{w(u)} du \right) = \int_{\mathbf{R}^n} f(u) \frac{\partial \overline{w(u)}}{\partial x_i} du \quad (16)$$

where the differentiation is done under the integral sign with respect to the wavelet co-ordinates x_i , not the independent variable u of the wavelet functions. This method can be more efficient than using approximation by finite differences, although sometimes analytic differentiation of the integrand is messy.

Local search methods based on partial derivatives tend to operate more efficiently if the variations in the SAWT

in each direction are of similar magnitude. For example, the conventional wavelet transform applied to 1-dimensional signals is a function of two variables: position and scale. At small scales the variability of this function is much greater with respect to position than scale and at large scales the opposite is true. This asymmetric behaviour results in narrow ridges and valleys in the cost function and reduces the efficiency of search methods, which navigate these irregularities using expensive calculations of the cost function. The Riemannian metric, defined next, quantifies this asymmetry, the latter being co-ordinate dependent, so we choose local co-ordinates for which this metric is spherically symmetric.

4.2. Wavelet space metrics

We define the following metric $d: W \times W \rightarrow [0,1]$:

$$d(w_1, w_2) = 1 - \langle w_1, w_2 \rangle, \quad (17)$$

for all $w_1, w_2 \in W$. We show below that this choice of metric describes the variability of the SAWT, both statistically in white noise and deterministically. First we consider the deterministic case. Let w_1 be a wavelet in the wavelet space. We define a metric ball $B(w_1, \epsilon)$ centred at this wavelet to be given by:

$$B(w_1, \epsilon) = \{w_2 \in W : d(w_1, w_2) < \epsilon\}. \quad (18)$$

Because $\langle f, w \rangle = \langle f, e^{i\theta} w \rangle$ for all θ , we can choose θ for any particular choice of w_1 and w_2 so that

$$\langle w_1, w_2 \rangle = \langle w_1, w_2 \rangle \cdot e^{i\theta}. \quad (19)$$

Using the Cauchy-Schwarz inequality it then follows that:

$$\begin{aligned} \|T_f(w_1) - T_f(w_2)\| &= \|\langle f, e^{-i\theta} w_1 \rangle - \langle f, w_2 \rangle\| \\ &\leq \|\langle f, e^{-i\theta} w_1 - w_2 \rangle\| \\ &\leq \|f\|_2 \|e^{-i\theta} w_1 - w_2\|_2 \\ &= \|f\|_2 \sqrt{2 - 2 \operatorname{Re}(\langle e^{-i\theta} w_1, w_2 \rangle)} \\ &= \|f\|_2 \sqrt{2 - 2 \langle w_1, w_2 \rangle} \\ &\leq \sqrt{2\epsilon} \|f\|_2. \end{aligned} \quad (20)$$

Not only does (20) give an upper bound on the variability of the SAWT within a metric ball, it also shows that if the wavelets in the wavelet space vary smoothly with respect to a chart or co-ordinate systems, then their SAWT values will also vary smoothly.

We now take the SAWT to be real-valued and consider the statistical case. Let n be a Gaussian white noise random process. The difference in the SAWT values of

the wavelets w_1 and $w_2 \in B(w_1, \epsilon)$ is given by $\langle n, w_1 - w_2 \rangle$ which has a mean of zero and a standard deviation proportional to the L_2 norm of $w_1 - w_2$. The latter is given by

$$\|w_1 - w_2\|_2 = \sqrt{2 - 2 \langle w_1, w_2 \rangle} \quad (21)$$

which for $1 - \epsilon > 0$ gives $\|w_1 - w_2\|_2 \leq \sqrt{2\epsilon}$.

Our choice of metric is in general expensive to compute, involving the construction of the wavelets from their co-ordinates and then integration over their common support to evaluate the inner product. We therefore employ a less expensive metric, which has similar properties. We define a Riemannian metric [13] $R_{ab} dx^a dx^b$ at the wavelet w to be half the second covariant tensor derivative of metric (17) where the first wavelet is fixed at w , and differentiation is with respect to the second wavelet. In practice, this metric is a second order local approximation of (17) because the first partial derivatives are zero, and thus

$$R_{ab} dx^a dx^b = 1 - \langle w_x, w_{x+dx} \rangle + o(dx^2) \quad (22)$$

4.3. Wavelet grids

Before explaining how to calculate R_{ab} efficiently, we show how it can be used to construct global search grids. Suppose that a grid consisting of wavelets $\{v_k\}$ in the sub-manifold $V \subset W$ is chosen whose metric balls $\{B(v_k, \epsilon)\}$ cover that part of the sub-manifold to be searched. From (20), we know that by choosing $\epsilon > 0$ small enough, we can conduct a global search for SAWT values above a given threshold, T say, by evaluating the SAWT once in each of the metric balls $\{B(v_k, \epsilon)\}$ and testing against a slightly lower threshold of $T(1 - \sqrt{2\epsilon})$. The Riemannian metric determines to first order in local co-ordinates the size of the metric balls $\{B(v_k, \epsilon)\}$, and thus the resolution required for the initial grid used in the global search. It is therefore sufficient to construct a grid whose spacings dx are to second order constrained by:

$$R_{ab} dx^a dx^b < \epsilon. \quad (23)$$

4.4. Local co-ordinates and group actions

In general the Hessian (matrix of second derivatives) and thus the Riemannian tensor are expensive to compute, being different at each point in the wavelet space, but by augmenting the manifold with additional group structure, we can derive the Hessian at any point in the wavelet space from the Hessian at a single point. The additional structure comes by considering group actions on wavelets. We now consider wavelet spaces which are the

orbit of a single (mother) wavelet w_0 (analogous to the mother wavelets used in Meyer [1]),

$$W = G \cdot w_0 = \{g \cdot w_0 : g \in G\} \quad (24)$$

under a group G . All the examples of wavelet spaces considered in Section 2 have this group structure: Fourier transforms involve frequency shifts, wavelet transforms involve translation and dilation, and Hough transforms involve translation and rotation. In order to maintain the normalisation of the wavelets, we also require that G is isometric, that is it preserves inner products. As we now show, the group action provides us with a means of deriving a local co-ordinate system at each wavelet w in the wavelet space in which the Riemannian tensor at w is identical to the Riemannian tensor at the mother wavelet.

We define the function $p: G \rightarrow W$ from the group to the wavelet space by

$$p: g \mapsto g \cdot w_0. \quad (25)$$

Because $G \cdot w_0$ spans the wavelet space, we can choose a canonical function $q: W \rightarrow G$ from the wavelet space to the group to be a right inverse of p , i.e. $p \circ q$ is the identity map on W . We then have

$$q: w \mapsto g \quad \text{where} \quad g \cdot w_0 = w, \quad (26)$$

i.e., q is chosen to canonically map the wavelet w to a group element whose action on the mother wavelet produces w .

For any two wavelets $w_1, w_2 \in W$ we consider the following combination:

$$w = p(q(w_1)q(w_2)), \quad (27)$$

which allows us to define an isometric transformation between local co-ordinate systems which preserves the co-ordinates of the Riemannian tensor. Let $c_i: W \rightarrow \mathbf{R}^n$ be a local co-ordinate system for which the Riemannian tensor is equal to R at the mother wavelet w_0 and let $w' \in W$ be any other wavelet, then the local chart which makes the co-ordinates of the Riemannian tensor equal to R at w' is given by

$$c: w \mapsto c_i \circ p(q(w')q(w)). \quad (28)$$

To see that this transformation is isometric, let $w_x = c^{-1}(x)$ and $w_y = c_i^{-1}(y)$ for the two charts given above, and choose any two wavelets $w_1, w_2 \in W$. From equations (25) and (28) and the isometric nature of G , it follows that

$$\begin{aligned} \langle w_{x_1}, w_{x_2} \rangle &= \langle p(q(w')q(w_{y_1})), p(q(w')q(w_{y_2})) \rangle \\ &= \langle q(w')q(w_{y_1})w_0, q(w')q(w_{y_2})w_0 \rangle \quad (29) \\ &= \langle q(w_{y_1})w_0, q(w_{y_2})w_0 \rangle \\ &= \langle w_{y_1}, w_{y_2} \rangle. \end{aligned}$$

We thus need only evaluate the Riemannian tensor from the Hessian matrix at one point in the wavelet space, and use equation (28) to derive this tensor at any other point. If we design a local co-ordinate system which has "good behaviour" at any point, then this co-ordinate transformation provides us with a means to transport this behaviour to any other wavelet in the wavelet space. In our case (Section 2.4) we would like the Riemannian metric to be spherically symmetric, which is done by designing a local chart for which the Riemannian tensor R_{ab} is the identity matrix. It is always possible to do this because the tensor is symmetric positive definite, so an orthogonal set of eigenvectors can be found and normalised so that the diagonal elements of this matrix are unity. If R_{ab} has the following diagonal decomposition

$$R = UDU^T \quad (30)$$

with respect to the chart c_2 where D is diagonal and U is unitary, then the required local co-ordinate system c_1 is given by solving

$$c_2 = D^{1/2} U c_1(w). \quad (31)$$

4.5. 1D example: Acoustic signals

To characterise acoustic signals, we consider the Hilbert space F_1 of 1-dimensional signals, a mother wavelet $w_0 \in F_1$ and a 4-dimensional group G of signal transformations defined by

$$[(p, s, a, b) \cdot f](t) = \frac{1}{\sqrt{s}} f\left(\frac{t-p}{s}\right) e^{i\left(a\left(\frac{t-p}{s}\right) + b\left(\frac{t-p}{s}\right)^2\right)} \quad (32)$$

and parameterised by translation p , dilation s , frequency shift a and chirp shift b , where a and b are normalised with respect to dilation. Examples of these wavelets are shown in Fig. 1. Group multiplication is defined to be the composition of these mappings, and in the above co-ordinate system is given by:

$$\begin{pmatrix} p_1 \\ s_1 \\ a_1 \\ b_1 \end{pmatrix} \times \begin{pmatrix} p_2 \\ s_2 \\ a_2 \\ b_2 \end{pmatrix} = \begin{pmatrix} p_1 + p_2 s_1 \\ s_1 s_2 \\ (a_1 + 2b_1 p_2) s_2 + a_2 \\ b_1 s_2^2 + b_2 \end{pmatrix}. \quad (33)$$

The wavelet space W is defined to be the orbit of w_0 when acted upon by G , and is thus the set of translations, dilations, frequency and chirp shifts of the mother wavelet w_0 . These functions are sometimes called chirplets. In this particular case there is a one-to-one

correspondence between G and W , so it is possible to identify each wavelet with its group action on the mother wavelet, and use the same co-ordinate system for both.

This chirp wavelet space $W = G \cdot w_0$ is a superset of the dictionaries used in the Fourier, Wavelet and Gabor Transforms. Any section of the SAWT restricted to constant a and b is a conventional Wavelet Transform parameterised by position p and scale s . The parameters a and b determine the shape of the mother wavelet in each of these sub-manifolds.

The initial search for local maxima in the SAWT is restricted to the 3-dimensional sub-manifold where the chirp angle b is zero. This subspace can be parameterised by position, scale and ordinary frequency $a' = a/s$. The Riemannian tensor with respect to these co-ordinates is given by:

$$R = \begin{pmatrix} As^{-2} & 0 & 0 \\ 0 & Bs^{-2} & 0 \\ 0 & 0 & Cs^2 \end{pmatrix} \quad (34)$$

where A, B, C are constants which depend on the mother wavelet. An approximately Gaussian envelope was chosen for the mother wavelet which is of compact support and continuously second differentiable, ensuring that the SAWT is equally smooth. We require a search grid for which the Riemannian metric is as even as possible for an even coverage of the wavelet space. Equations (20) and (23) imply that the spacings dp and ds should be proportional to scale s and the spacing da' should be *inversely* proportional to scale where the constants of proportionality are given by the reciprocals of the square roots of A, B and C . A grid meeting these requirements is obtained by dividing the sub-manifold into sections of constant scale, and in each section defining a regular rectangular grid where the spacing in position is proportional to scale and that of frequency inversely proportional to scale, so that *different grids* are used at different scales. The scales themselves need to be a geometric sequence to generate the correct spacings ds .

The SAWT at each fixed scale, essentially a Gabor Transform, is computed efficiently as a set of windowed FFT's. With a Gaussian envelope a low resolution in scale is required, with powers of 2 usually being sufficient, in which case a scale range of 1000:1 is covered with only 10 scale intervals and 11 Gabor Transforms. The complex exponential wavelets are better suited to characterising an analytic signal, where there is no energy at negative frequencies in the Fourier Transform, than a real-valued signal, which has equal energy at positive and negative frequencies. For real-valued signals the Hilbert Transform is applied first to create an analytic signal, where sinusoids are converted to complex exponentials.

The use of complex exponentials in the wavelet has a number of advantages over real sinusoids. The phase of the sinusoid is controlled by wavelet amplitude and so is not required as an additional parameter in the search space, making the search more efficient. A single wavelet in the signal data results in a single local maximum in the SAWT in the complex case, whereas there are usually many local maxima in the real case, caused by the sinusoid in the signal being alternatively in and out of phase with the filter. The lack of "side-lobes" in the complex case greatly reduces the redundancy of the wavelet characterisation. The smoother SAWT also enables the local search for local maxima to work more efficiently. The group G is also easier to define in the complex case.

Fig. 2 demonstrates the chirp wavelet representation on an example of an underwater acoustic signal attributed to dolphins. Fig 2a shows a spectrogram (Gabor Transform) of the signal data, showing the variation of signal energy with time and frequency, in which several coherent structures can be perceived against a noisy background. Fig 2b shows the corresponding spectrogram of a signal reconstruction from a chirp wavelet decomposition, which has extracted most of the interesting structure. Only 15 chirp wavelets are used in the characterisation, demonstrating the economy of the method. The wavelets cover a wide range of frequency, scale and chirp angle, and it would not be possible to characterise this data as efficiently with more limited dictionaries such as the Gabor or Wavelet dictionaries.

4.6. 2D example: Sea surface imagery

To characterise 2D imagery, we consider the Hilbert space F_2 of images, and we take our wavelet space W to consist of the Gaussian ellipses:

$$w_{(x,y,s,\theta,a)}(u,v) = \exp\left(-\left(\frac{U^2}{s^2} + \frac{V^2}{a_0 a^2 s^2}\right)\right)$$

where

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} u-x \\ v-y \end{pmatrix} \quad (35)$$

Our wavelets thus have a chart to \mathbf{R}^5 , where (x, y) is the position of the centre of the wavelet relative to some origin, s is the scale of the wavelet, a is its aspect ratio and θ is its orientation with respect to the y -axis. We choose the constant a_0 to be larger than unity to remove a co-ordinate system singularity which occurs because of the lack of effect rotation has upon wavelets with aspect ratio one.

The natural choice of image transformations for this wavelet space would contain translation (for position), dilation (for scale), rotation (for orientation) and stretching (for aspect ratio); however, the smallest group

containing these transformations is the affine group, which is six-dimensional, and includes skew transformations. We parameterise this group G as $\{(A, b)\}$ where A is a 2×2 matrix composed of rotation, dilation, stretching and skew transformations, and where b is a column vector representing translation. We define group multiplication to be given by

$$(A_1, b_1)(A_2, b_2) = (A_1 A_2, A_1 b_2 + b_1), \quad (36)$$

and the inner product preserving group action on F_2 to be given by

$$[(A, b)w](u) = \frac{1}{\sqrt{\det(A)}} w(A^{-1}(u - b)), \quad (37)$$

for all $w \in F_2$.

Unlike the previous example, the parameter spaces of the wavelet space and the group can now no longer be directly identified with one another by judicious choice of charts. Indeed, while the group G can only be charted to 6-dimensional Euclidean space, the wavelet space W has been charted to 5-dimensional space. This means that the co-ordinate versions of the functions $p: G \rightarrow W$ and $q: W \rightarrow G$ used in the chart transformation (28) are no longer trivial. We could have removed this difficulty by choosing a chart mapping the wavelet space to \mathbf{R}^6 , but this would have increased the dimension and thus the difficulty of searching for local maxima in the SAWT.

The initial search for local maxima is restricted to a 4-dimensional sub-manifold with the bar aspect ratio fixed to 3, that is fairly broad bars. In most wavelet spaces four dimensions is too computationally demanding, but for broad bars the SAWT is insensitive to orientation and scale, so only a small number of combinations of these parameters is required in the search grid. Where there are long narrow objects, such as the ship wakes in Fig. 2, the local search for unconstrained local maxima in the SAWT extends the broad bars along these structures. The SAWT is much more sensitive to orientation for narrow bars than broad bars, but this and other dependencies on location in the wavelet space are removed by the local co-ordinate systems defined in Section 4.4.

Fig. 3 illustrates the use of bar wavelets to detect and characterise ships and their wakes from electro-optic imagery of the sea surface. Both extended structures, the wakes, and a more compact structure, the ship, are characterised efficiently by this method. The ship's course, including speed and direction, can be accurately inferred from the wavelet parameters, including the change in course half way along the wake.

5. Removing redundancy and reconstruction

In general the wavelets $\{w_k\}$ detected as local maxima in the SAWT are not mutually orthogonal. This has two undesirable effects:

- More wavelets than necessary may be required to represent a given signal or image (or the foreground) with a given level of accuracy.
- Optimal reconstruction is no longer obtained by the linear superposition of single wavelet reconstructions.

Traditionally these problems are avoided by designing the wavelet basis functions to be orthogonal, for example using the wavelets of Daubechies [2]. This approach is fundamentally incompatible with optimisation in a differentiable manifold of dictionary functions because there is an open neighbourhood $B(w_1, \epsilon)$ of each wavelet w_1 in which $\langle w_1, w_2 \rangle$ is greater than zero. The inherent flexibility of the local search method, where local approximations to the signal or image are optimised, and where there is potential for fine adjustments to the wavelet parameters, is incompatible with the necessary constraints required for orthogonality.

5.1. Gram-Schmidt orthogonalisation

We therefore remove the constraint of orthogonality in the wavelet decomposition and instead remove redundancy either by excluding wavelets or by adjusting the coefficients in the wavelet series expansion, initially given by the SAWT. The process is based on Gram-Schmidt orthogonalisation, where a temporary orthonormal basis $\{e_n\}$ is constructed as *linear combinations* of wavelets $\{w_k\}$ extracted as local maxima in the SAWT. The new basis functions are given by:

$$E_n = w_n - \sum_{k=1}^{n-1} \langle w_n, e_k \rangle e_k, \quad e_n = \frac{E_n}{\|E_n\|} \quad (38)$$

where f is the signal or image data. Instead of calculating the signal or image functions $\{E_n\}$ and $\{e_n\}$ directly, we calculate recursively the coefficients of the linear combinations of $\{w_k\}$ from which they are constructed using equation (38).

Unlike ordinary Gram-Schmidt orthogonalisation, the order in which the wavelets $\{w_k\}$ are chosen is not fixed but is adapted to maximise the rate of decrease of the residual sum-of-squares error. This is achieved by choosing w_n to maximise the inner product:

$$I = |\langle f, e_n \rangle| \quad (39)$$

which in turn minimises the residual sum-of-squares error given by:

$$\left\| f - \sum_{k=1}^n \langle f, e_k \rangle e_k \right\|^2 = \left\| f - \sum_{k=1}^{n-1} \langle f, e_k \rangle e_k \right\|^2 - I^2. \quad (40)$$

The process stops when the values of I for all the remaining wavelets $\{w_k'\}$ are below the amplitude threshold for unusual objects, that is when none of the wavelets contributes sufficient information about the signal data f which is *independent* of the wavelets already included in the expansion. The value I for each remaining candidate wavelet w_n is derived from the inner products $\langle f, w_n \rangle$ and $\langle w_m, w_n \rangle$, where the former is given by the SAWT (4) and the latter are independent of the signal and depend solely on the wavelet parameters. The following recurrence relations are used to derive I :

$$\langle f, e_n \rangle = \frac{\langle f, w_n \rangle - \sum_{k=1}^{n-1} \langle f, e_k \rangle \langle e_k, w_n \rangle}{\sqrt{1 - \sum_{k=1}^{n-1} \langle e_k, w_n \rangle^2}} \quad (41)$$

$$\langle w_n, e_i \rangle = \frac{\langle w_n, w_i \rangle - \sum_{k=1}^{i-1} \langle w_n, e_k \rangle \langle e_k, w_i \rangle}{\sqrt{1 - \sum_{k=1}^{i-1} \langle e_k, w_n \rangle^2}} \quad (42)$$

5.2. Wavelet subtraction

The above wavelet expansion is generally a good approximation to the signal or image, but is not usually an exact representation, and therefore not in general invertible. In cases in which the background is not of interest and its rejection is required, the incompleteness of the representation is usually not a drawback as the residual error is dominated by information which is not of interest. In situations where the accuracy of the wavelet reconstruction is not sufficient, however, a process called *wavelet subtraction* is employed which adds further wavelets to the expansion to achieve an arbitrarily small residual.

The wavelet subtraction process starts with a wavelet expansion from local maxima in the original SAWT with redundancy removed using Gram-Schmidt orthogonalisation. The signal reconstruction from this expansion is then subtracted from the original signal to leave the residual f_r . A new wavelet expansion is then computed for the residual f_r , in which local maxima of a new SAWT are computed. The new wavelets are then added to the original set and redundancy removal is repeated for this union. This process is repeated until the required level of accuracy is obtained.

6. Sensor invariant representation

We now consider how to represent foreground information, with the aid of the wavelet features extracted using shape adaptive wavelet analysis, in a manner which depends as little as possible on the relationship between the objects being viewed and the sensor. A frequently encountered example of this [14] is the dependence of the relative 3D position and orientation of the object and an imaging sensor. This example will be considered in this paper to illustrate the method. Another example is the effect of range and relative velocity on acoustic data, where range affects amplitude and timing of acoustic transients, and changes in velocity cause a Doppler frequency shift.

6.1. Sensor-related transformations

In general we represent sensor-related effects as a set of possible transformations G acting on the signal or image data. The objective is to derive from the wavelet features a set of numbers which does not change under the action of any transformation in G , but from which all other attributes of the ensemble of features can be determined uniquely. If this objective can be met, these numbers can be used to classify objects in a sensor-invariant manner. Given a set of image features $\{w_i\}_1^n$, that represent some real-world object, and a transformation $g \in G$, we can form another set of features $\{g.w_i\}_1^n$, where $g.w_i$ denotes the action of the transformation g on the wavelet feature w_i . (We encountered a similar situation in Section 4.4, where the wavelet space itself was generated as the orbit of a group of transformations acting on the mother wavelet.) These feature sets are considered equivalent because they may represent the same target from a different viewpoint.

6.2. Canonical forms and alignment functions

Wavelet sets which are linked in this way by a transformation in G form an equivalence class, and our approach is to define a unique member for each such equivalence class, which we call a canonical form. The parameters of the image features in the canonical form are the sensor-invariant numbers we require. Our aim is to derive this canonical form from any wavelet set belonging to the same equivalence class, because then we can characterise the real-world object in the same way, no matter what the sensor configuration is when the object is detected. We do this by finding an element g in G which will transform any given feature set $\{w_i\}_1^n$ to the canonical form $\{g.w_i\}_1^n = \{c_i\}_1^n$. The transformation g will depend on the set $\{w_i\}_1^n$ but will transform equivalent feature sets to the same canonical form $\{c_i\}_1^n$. To find g , we introduce an alignment function m that

returns a vector of measurements for a given set of features,

$$m: \{w_1, \dots, w_n\} \mapsto x = (x_1, \dots, x_N), \quad (43)$$

where N is the dimension of the transformation group G . The purpose of the alignment function is to provide enough constraints on the feature set to fix g . First we take an alignment measurement of the canonical form,

$$m(c_1, \dots, c_n) = m_0. \quad (44)$$

We now find the transformation g that satisfies

$$m(g.w_1, \dots, g.w_n) = m_0. \quad (45)$$

This involves solving a set of simultaneous equations, and we choose the alignment function such these equations are non-singular, and ideally easy to solve, for example by being linear. The transformation g will then also satisfy

$$\{g.w_1, \dots, g.w_n\} = \{c_1, \dots, c_n\}. \quad (46)$$

To see that the canonical forms are invariant under group transformation, suppose we have another set of features from the same equivalence class

$$\{h.w_1, \dots, h.w_n\}. \quad (47)$$

The transformation which ensures that the alignment function applied to these features matches the canonical alignment measurement m_0 is then gh^{-1} . When applied to the new features we once again produce the canonical set

$$\begin{aligned} \{gh^{-1}h.w_1, \dots, gh^{-1}h.w_n\} &= \{g.w_1, \dots, g.w_n\} \\ &= \{c_1, \dots, c_n\} \end{aligned} \quad (48)$$

6.3. Example alignment functions

We now consider three examples of alignment functions, each requiring a choice of features from the set comprising the foreground, which we call anchor features. The choice of anchor feature(s) leaves an ambiguity in the canonical form; however, for a small number of anchor features and a moderate number of foreground features, the number of permutations is manageable. In the examples given in Section 7.2, the number of wavelets is of the order 5 and the number of anchor wavelets no more than 3, so the number of permutations typically 20 or less.

The first example is applicable to any wavelet decomposition for which the sensor-related transformations are the same as the group generating the wavelet space from the mother wavelet (Section 4.4). In this case we consider the alignment function which returns the co-ordinates of one of the features in the set:

$$f(w_1, \dots, w_n) = w'_1 \quad (49)$$

where w'_1 are the co-ordinates of the chosen wavelet w_1 for some co-ordinate system. We choose the anchor wavelet to be the mother wavelet. The canonical alignment measurement then becomes $c'_1 = w'_0$. We now make use of the mapping q defined in equation (26) of Section 4.4 which associates group actions with wavelets. In the case where there is a one-to-one correspondence between the wavelet space and the group generating it, q is bijective, and the group transformation which satisfies equation (45) for our particular alignment function is $q(w_1)^{-1}$. The canonical wavelets are then produced through application of $q(w_1)^{-1}$ to each wavelet in the set

$$\{q(w_1)^{-1}.w_1, \dots, q(w_1)^{-1}.w_n\}. \quad (50)$$

The values in Equation (50) represent the relationships between each image feature and the anchor feature, and these relationships, being group quotients, are invariant under the action of the group.

There are many examples of wavelet group actions where q is bijective, for example any group spanned by any combination of translation, dilation, and frequency shift, including the wavelet and Gabor transforms. The above technique applied to features generated by the Gabor transform is therefore useful for acoustic transient recognition, being invariant to time of arrival, range and Doppler shifts.

The next two examples concern aspect-invariant recognition of objects in 3D space viewed at long range, where we follow the commonly adopted approach of modelling changes in viewing aspect as affine transformations in the image of the scene [14]. We use the wavelet space generated from the affine group acting on a Gaussian radial basis function (Section 7.2). Unfortunately the mapping q is not bijective, as there affine group has 6 dimensions and the wavelet space only 5, so the alignment measurement is insufficient to uniquely specify the transformation satisfying equation (45). Instead, we require alignment functions which return a 6-element vector, as in the remaining two examples.

The second alignment function returns the positions of three anchor wavelets within the set,

$$f(w_1, \dots, w_n) = (x_{w_1}, y_{w_1}, x_{w_2}, y_{w_2}, x_{w_3}, y_{w_3}). \quad (51)$$

With this particular case it is usually possible to identify uniquely the transformation satisfying equation (45). Suppose that the canonical alignment measurement is $(x_{c_1}, y_{c_1}, x_{c_2}, y_{c_2}, x_{c_3}, y_{c_3})$, and the affine transformation to be determined is given by (a, b, c, d, e, f) , where

$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is the composite matrix of scaling, aspect

change, skewness and rotation, and $\begin{pmatrix} e \\ f \end{pmatrix}$ is the translation offset. The transformation is then determined by the linear system

$$\begin{pmatrix} x_{w_1} & y_{w_1} & 0 & 0 & 1 & 0 \\ 0 & 0 & x_{w_1} & y_{w_1} & 0 & 1 \\ x_{w_2} & y_{w_2} & 0 & 0 & 1 & 0 \\ 0 & 0 & x_{w_2} & y_{w_2} & 0 & 1 \\ x_{w_3} & y_{w_3} & 0 & 0 & 1 & 0 \\ 0 & 0 & x_{w_3} & y_{w_3} & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} = \begin{pmatrix} x_{c_1} \\ y_{c_1} \\ x_{c_2} \\ y_{c_2} \\ x_{c_3} \\ y_{c_3} \end{pmatrix} \quad (52)$$

If the feature centres are not collinear, the system is non-singular and therefore has a unique solution. Collinearity can be detected prior to the derivation of the canonical form and such cases rejected at an earlier stage in the recognition process. Anchor features in known target types can be chosen to be strongly non-collinear, unless the target has a particularly simple structure, in which case it should be easy to recognise anyway.

The third alignment function returns the positions and orientations of two anchor wavelets

$$f(w_1, \dots, w_n) = (x_{w_1}, y_{w_1}, \theta_{w_1}, x_{w_2}, y_{w_2}, \theta_{w_2}) \quad (53)$$

The transformation satisfying equation (45) is determined in a similar manner to the previous example. First, three sets of positions are produced from the measurement, where the first two correspond to the positions of the anchor wavelets, and the third set is the position of the intersection of the major axes of the two anchor wavelets. Once more there are singularities within the alignment function, in this case when the two anchor wavelets are parallel. As in the previous case in most cases judicious choices of image features in the canonical form should avoid this type of singularity.

6.4. Examples of 3D Aspect-Invariant recognition

We now present results for example aircraft appearing in infrared imagery. We first illustrate the uniqueness of the canonical forms for different aircraft, demonstrating their potential for class discrimination. Fig. 4 shows infrared images of the Sukhoi-30 fighter, the B2 Stealth Bomber and a transporter plane. Fig. 5 shows a wavelet approximation of each of the aircraft. Notice how the Sukhoi-30 and B2 have been approximated by wavelets along the edges of the aircraft outlines, whereas the transporter plane has had wavelets fitted to the main fuselage and wings. The canonical forms for these aircraft are shown in Fig. 6, and their dissimilarity clearly demonstrates the potential ease with which different types of aircraft can be distinguished.

In figures 7 and 8 we demonstrate the invariance of the canonical form to different viewing aspects. Fig. 7 shows four different views of the transporter plane. In Fig. 8 we show different canonical forms of these images superimposed to illustrate their similarity. The three wavelets chosen to be the anchors are the fuselage, the tail wing and one of the engines. In Fig. 8(a) we compare the canonical forms extracted from Fig 7(a) and Fig. 7(b), and in Fig. 8(b) we compare the canonical forms extracted from Fig 7(c) and Fig. 7(d). Most of the image feature parameters have changed very little, indicating that a template classifier would have little difficulty in recognising the similarity of the aircraft from this type of representation. This similarity has been observed for real data, in spite of the limitations of the affine transformation model, which does not take account of pixel aliasing, occlusion, non-planarity and sensor distortion.

In figures 9 and 10 we demonstrate the stability of the canonical representations of the transporter plane to the addition of simulated noise to investigate likely performance caused either by a degraded sensor, or atmospheric obscuration. There is already some noise in the original image data, but this data is of higher quality than could be encountered in poor weather. Each canonical form been generated using the second alignment function of Section 6.3, where the anchor wavelets are chosen to be the fuselage, the tail wing and one of the engines. Fig. 9(a) shows the image with Gaussian noise added and Fig 10(a) compares the canonical form generated by from this image with that of the original image. Figs. 9(b) and 10(b) show similar results with a lower SNR, where it can be seen that even in this case the canonical forms are still very similar.

7. References

1. Y.Meyer, *Wavelets: Algorithms and Applications*, SIAM, Philadelphia, 1993.
2. G.Strang and T.Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge press, Wellesley, MA, 1996.
3. R.R.Coifman and M.V.Wickerhauser, "Entropy-Based Algorithms for Best Basis Selection", *IEEE Trans. on Information Theory*, **38**, pp.713-718, 1992.
4. R.R.Coifman, Y.Meyer, S.Quake and M.V.Wickerhauser, "Signal processing and Compression with Wave Packets", *Numerical Algorithms Research Group*, New Haven, CT, Yale University, 1990. Available from anonymous internet ftp site at ceres.math.yale.edu ([130.132.23.22]) in `/pub/wavelets/cmqw.tex`
5. S.G.Mallat, "A Theory of Multi-Resolution Signal Decomposition: Wavelet Decomposition", *IEEE Trans. Patt. Anal. Mach. Intell.*, **1**, pp.674-693, 1989.

6. S.G.Mallat and W.L.Hwang, "Singularity Detection and Processing with Wavelets", *IEEE Trans. on Information Theory*, **37**, pp.617-643.
7. W.B.Davenport and W.L.Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, New York, 1958.
8. R.N.Bracewell, *The Fourier Transform and its applications*, McGraw-Hill, 1986.
9. J.Illingworth and J.Kittler, "A Survey of the Hough Transform", *Computer Vision, Graphics and Image Processing*, **44**, pp.87-116, 1988.
10. G.H.Watson and S.K.Watson, "Detection of unusual events in intermittent non-Gaussian images using multiresolution background models", *Optical Engineering*, **35**(11) pp.3159-3171, Nov 1996.
11. W.H.Press, *Numerical Recipes in C: the Art of Scientific Computing*, Cambridge University Press, 1992.
12. R.Fletcher, *Practical Methods of Optimisation: Unconstrained Optimisation*, Wiley-Interscience, New York, 1980.
13. C.T.J.Dodson and T.Potson, *Tensor Geometry*, Springer-Verlag, New York, 1991.
14. R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.

© British Crown copyright 1999. Published with the permission of the Defence Evaluation and Research Agency on behalf of the Controller of HMSO.

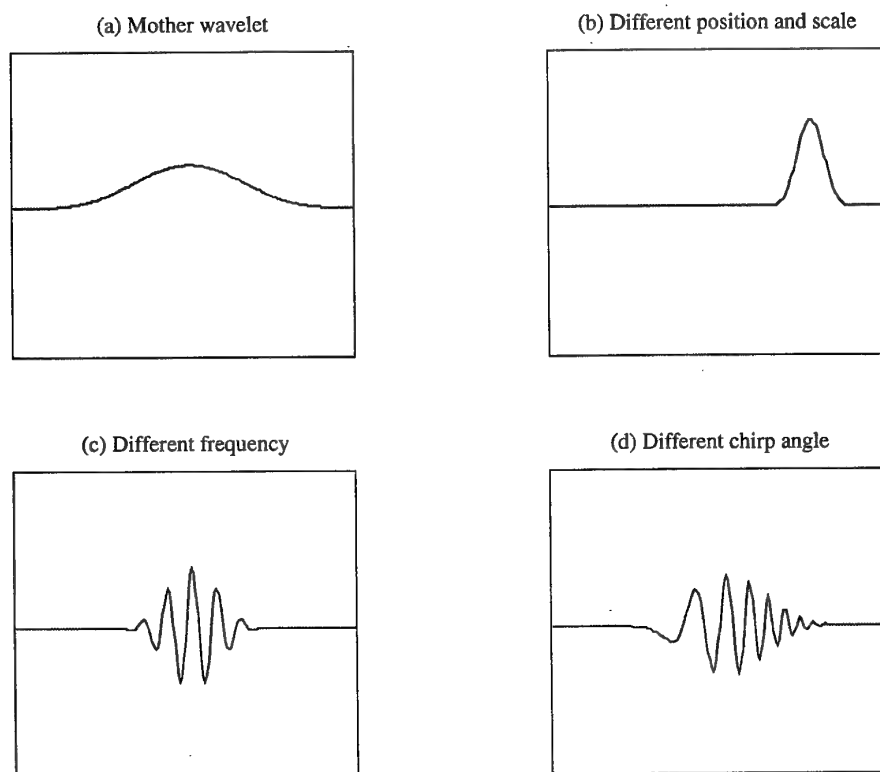


Fig. 1: Examples of 1D wavelets

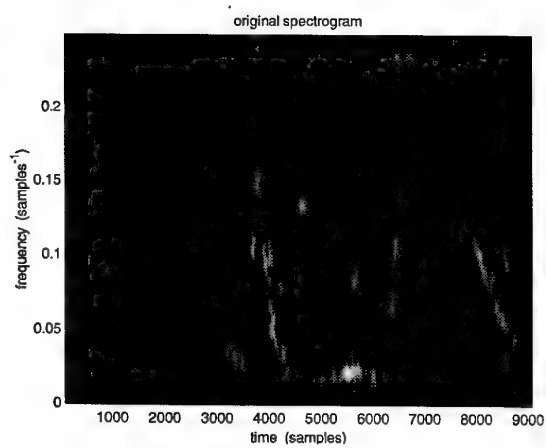


Fig. 2a: Spectrogram of underwater acoustic signal

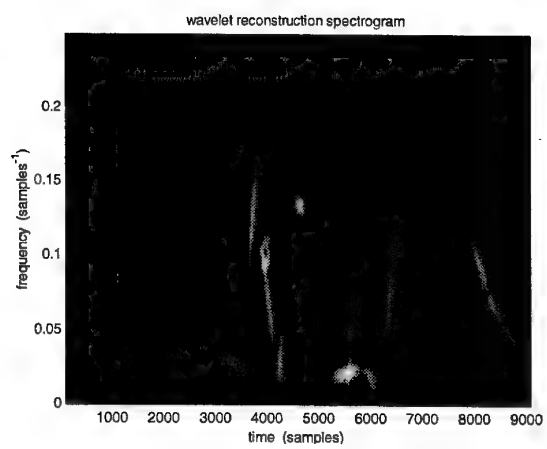


Fig. 2b: Chirp wavelet reconstruction of dolphin sounds

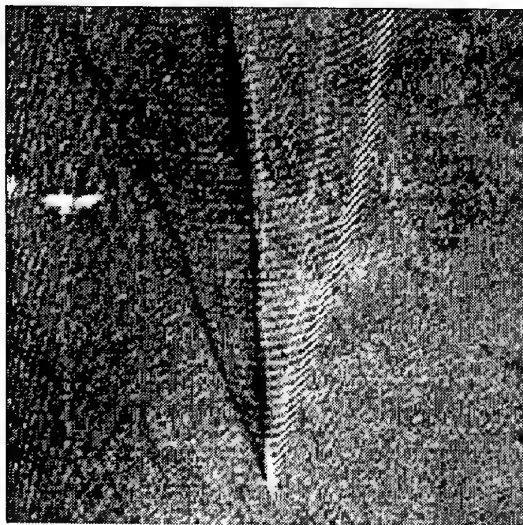


Fig. 3a: Satellite image of sea surface

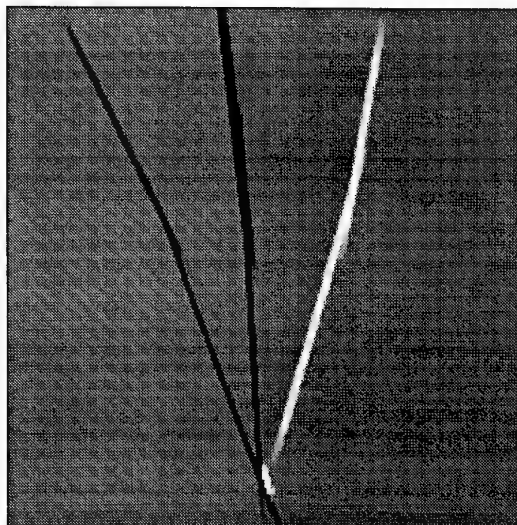
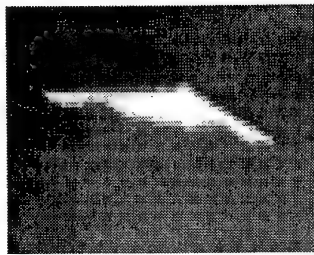


Fig 3b: Bar wavelet reconstruction

(a) Sukhoi-30



(b) B2 Bomber



(c) Transporter

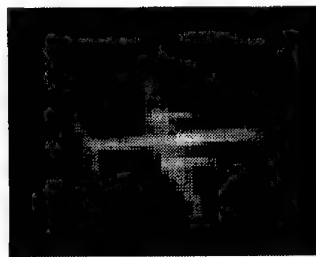
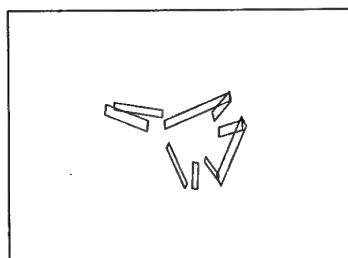
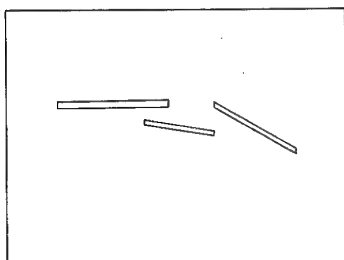


Fig. 4: Aircraft images (Infrared)

(a) Sukhoi-30



(b) B2 Bomber



(c) Transporter

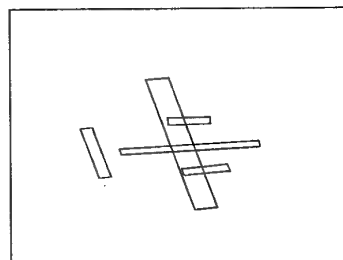
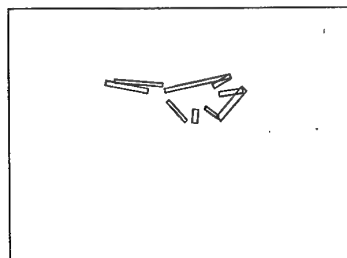
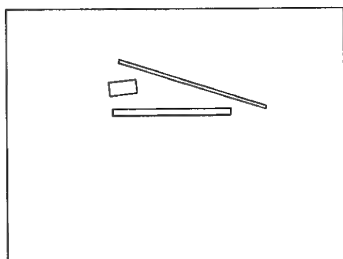


Fig. 5: Bar wavelet representations of aircraft images

(a) Sukhoi-30



(b) B2 Bomber



(c) Transporter

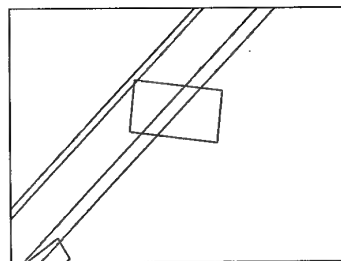


Fig. 6: Comparison of canonical forms from different aircrafts

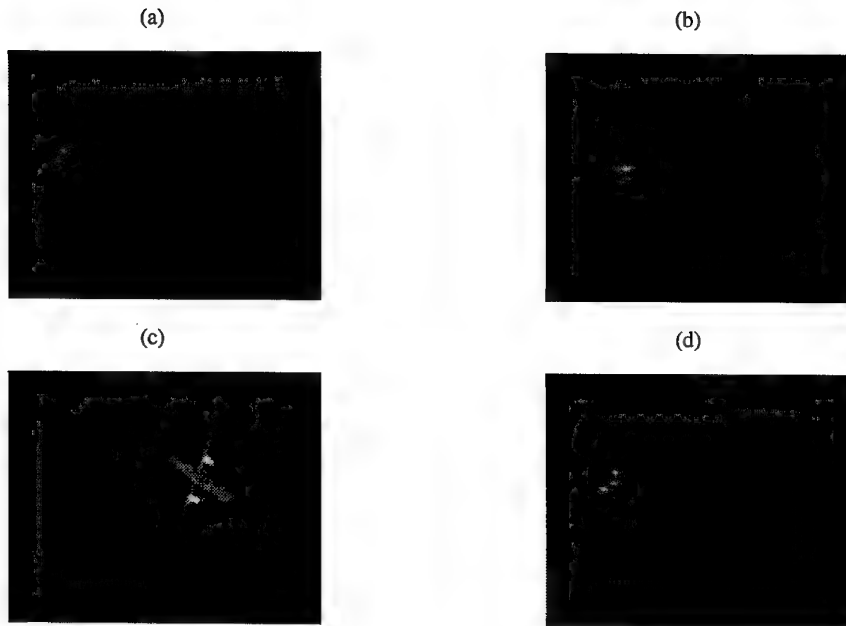
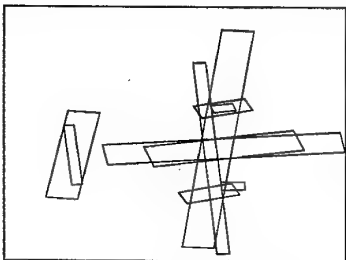


Fig. 7: Transporter images

(a) Comparison of Fig 7(a) and 7(b)



(b) Comparison of Fig 7(c) and 7(d)

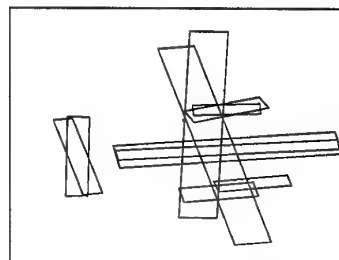
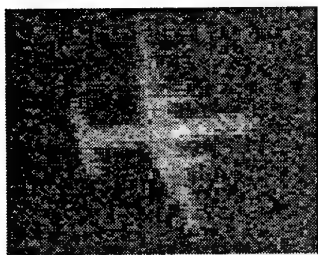


Fig. 8: Comparison of transporter canonical representations

(a) Medium noise



(b) High noise

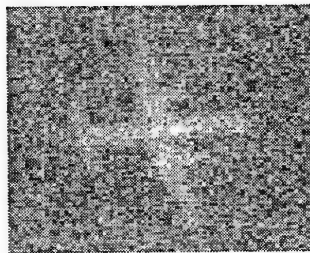
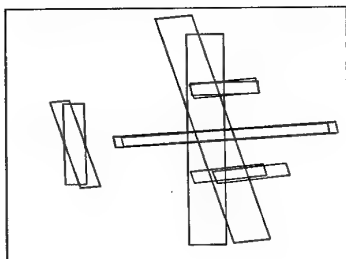


Fig. 9: Transporter image with extra gaussian noise

(a) Medium noise



(b) High noise

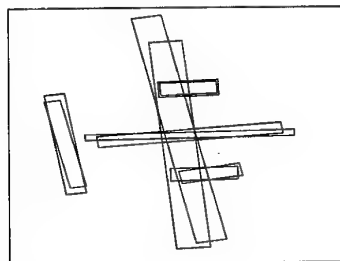


Fig. 10: Canonical forms with alignment function 2

Multirate filter banks and their use in communications systems[†]

Charles D. Creusere

Code 4T4400D

Naval Air Warfare Center

China Lake, CA 93555

email: creuserecd@navair.navy.mil

Abstract

Multirate filter banks are the fundamental building blocks of efficient wavelet and wavelet-packet implementations and are thus an important part of the current generation of image compression algorithms. In addition, multirate concepts also form the basis of local cosine transforms (LCTs) and lapped orthogonal transforms (LOTs) which are used in audio compression and noise removal applications. In this paper, we first discuss the fundamentals of multirate filter banks, both theory and implementation. Next, we focus on applications. In particular, we highlight wavelet-based image compression because research in this area has been very productive in recent years. We consider in particular embedded algorithms such as embedded zerotree wavelet (EZW) and set partitioning in hierarchical trees (SPIHT) because of the valuable capabilities they provide in a variety of military applications. Finally, we also discuss additional areas in which multirate filter banks play a roll including interference excision, signal scrambling, and code orthogonal frequency division multiplexing (COFDM) for data transmission.

1. Introduction

The theory of multirate signal processing has proven itself useful in a variety of applications over the last ten years. It is in communications, however, that multirate systems have thus far had the most impact. For example, the transforms used today in state-of-the-art image and video compression algorithms are themselves multirate systems. The discrete cosine transform (DCT) used in JPEG and MPEG can be viewed as a maximally decimated 2-dimensional 64-band multirate filter bank [1], [2]. Furthermore, the wavelet transforms used in the best still image compression algorithms (such as the upcoming JPEG 2000 standard) are themselves implemented as multirate filter banks [3]. Multirate filter banks are also at the center of all of the existing algorithms for wideband audio compression [4]. For example, the popular MPEG audio layers 1,

2, and 3 (the latter called MP3 by many) all use a 32-band cosine modulated multirate filter bank at their cores. Other algorithms such as Dolby AC-3 (now called Dolby Digital) and MPEG AAC (advanced audio coder) use lapped orthogonal transforms (LOTs) which are simply a special case of the general cosine modulated filter bank.

While multirate theory has had its greatest impact on compression systems, it has also influenced other areas of the communications field. In particular, multirate systems form the basis of time-frequency scrambling methods for secure voice communications [5] and they are now also being used to generate broadcast waveforms for code orthogonal frequency division multiple access (COFDM) [6]. Such systems are currently being used to broadcast digital radio in Europe and are likely to be used in the future for the broadcast of digital television as well.

This paper is organized as follows. Section 2 discuss the fundamentals of multirate signal processing, introducing the concepts of upsamplers and downsamplers. In Section 3, we introduce perfect reconstruction filter banks, and in Section 4 we discuss the application of such filter banks to image compression, wideband audio compression, noise removal, code orthogonal frequency division multiplexing, and signal scrambling. Conclusions are then presented in Section 5.

2. Fundamentals

To describe a digital filter, we use here three different representations [7]. In the time (or spatial) domain, a 1-dimensional (1D) digital filter is given by its impulse response $h(n)$ for $n \in (N_1, N_2)$. If N_1 and N_2 are finite, then $h(n)$ is a finite impulse response (FIR) filter; otherwise, it is an infinite impulse response (IIR) filter. We can also uniquely characterize this filter in the frequency domain by its Fourier transform as

$$H(e^{j\omega}) = \sum_n h(n)e^{-j\omega n} \quad (1)$$

or in the z-domain by

[†] Approved for public release, distribution unlimited

$$H(z) = \sum_n h(n)z^{-n}. \quad (2)$$

From a practical standpoint, we often deal with a discretized version of (1) called the discrete Fourier transform (DFT). The DFT calculates the coefficients of (1) at M sample points: i.e.,

$$H(k) = \sum_{n=0}^{M-1} h(n)e^{-j2\pi kn/M}. \quad (3)$$

The two basic operations of multirate digital signal processing are upsampling and downsampling as shown in Figs. 1 and 2 [8]. To upsample a digitized signal by a factor L , we add $L-1$ zeros between each input value to increase the sampling rate; to downsample the signal by M , we keep every M -th sample and discard the rest. The process of upsampling adds false high frequency content to the signal in the form of images while downsampling introduces aliasing into the output signal. Signal interpolation is accomplished by following the upsampler of Fig. 2 with a digital lowpass anti-imaging filter. To eliminate the aliasing introduced by downsampling, on the other hand, we must first filter the input signal $x(n)$ using an appropriate lowpass filter. Suffice it to say that even if aliasing is eliminated by prefiltering prior to downsampling, the lowpass filtering operation reduces the information content of the resulting signal. This is, of course, completely consistent with the theory of Nyquist sampling [7]. Finally, one notes from the example in Fig. 3 that any fractional sampling rate alteration can be achieved by using the basic integer building blocks shown in Figs. 1 and 2.

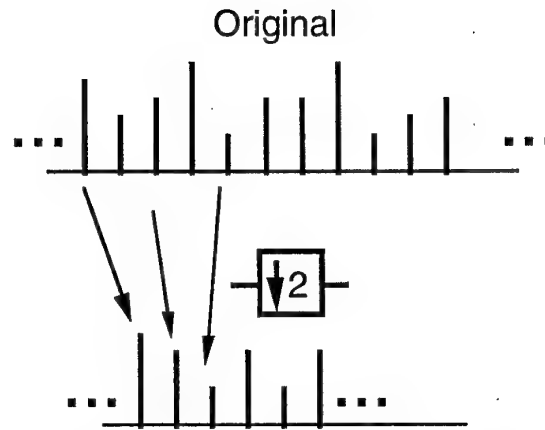


Figure 1: Downsampling by 2-- every other sample is retained.

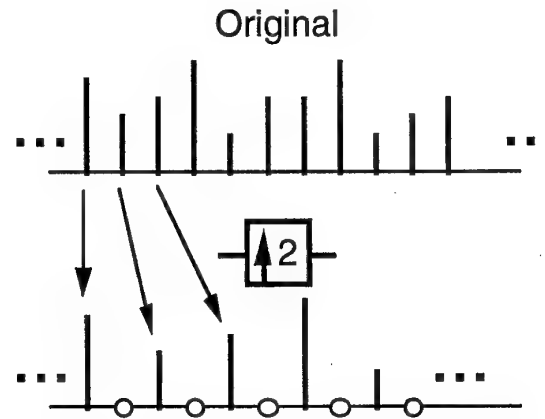


Figure 2: Upsampling by 2-- a zero is inserted between every other sample.

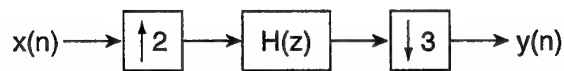


Figure 3: Decreases sampling rate by 2/3.

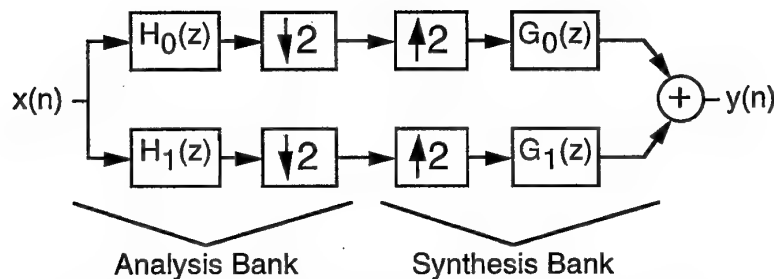


Figure 4: 2-band maximally decimated filter bank (e.g., wavelet).

3. Maximally decimated filter banks

3.1 2-band Systems

Using the multirate operators introduced in Section 2, we can now describe the maximally decimated filter bank. A 2-band analysis/synthesis system is shown in Fig. 4 where $H_0(z)$ is a lowpass filter and $H_1(z)$ is highpass. Note that because of the downsampling operations performed in the analysis filter bank, the sampling density in the transform or subband domain is exactly the same as it was in the original temporal domain. This is especially important for compression applications because each of the transform coefficients must be quantized and coded (i.e., converted into bits) for transmission. This property is also quite useful in other applications as well like pattern classification and signal scrambling [9], [5].

Using basic multirate identities and a little bit of algebraic manipulation on the system of Fig. 3, we find that

$$Y(z) = \frac{1}{2} [H_0(z)G_0(z) + H_1(z)G_1(z)] \cdot X(z) + \frac{1}{2} [H_0(-z)G_0(z) + H_1(-z)G_1(z)] \cdot X(-z) \quad (4)$$

[8]. The first term in (4) represents the linear time invariant (LTI) response of the combined analysis/synthesis system while the second represents the aliasing introduced into the system by downsampling. This aliasing can be completely canceled, however, if we select

$$G_0(z) = H_1(-z), \quad G_1(z) = -H_0(-z). \quad (5)$$

Substituting (5) into (4), we see that

$$Y(z) = T(z)X(z) \quad (6)$$

where

$$T(z) = \frac{1}{2} [H_0(z)H_1(-z) - H_1(z)H_0(-z)] \quad (7)$$

is the LTI transfer function of the system. In general, $T(z)$ may introduce both amplitude and phase distortion into the reconstructed signal. To achieve perfect reconstruction, $T(z)$ must have the form $c \cdot z^{-n_0}$ for constant c and integer n_0 .

Perfect reconstruction can be achieved with the appropriate choice of filters $H_0(z)$ and $H_1(z)$. In particular, assume that $H_0(z)$ is power symmetric. This implies that

$$\tilde{H}_0(z)H_0(z) + \tilde{H}_0(-z)H_0(-z) = 1 \quad (8)$$

where $\tilde{H}(z) \equiv H^*(z^{-1})$, indicating that the filter coefficients are first time reversed and then conjugated. To force $T(z)$ to have the desired form $c \cdot z^{-n_0}$ we can thus select

$$H_1(z) = z^{-N} \tilde{H}_0(-z). \quad (9)$$

For odd N , (6) reduces to $Y(z) = 0.5z^{-N}X(z)$ and perfect reconstruction is achieved! Substituting (9) into (5), the two synthesis filters are now given by

$$G_0(z) = z^{-N} \tilde{H}_0(z), \quad G_1(z) = z^{-N} \tilde{H}_1(z). \quad (10)$$

Note that all four filters in the system are completely determined by just one of them! The filter bank shown in Fig. 3 was first called the Conjugate Quadrature Filter Bank when described by Smith and Barnwell in [10]. It was not truly popularized, however, until Daubechies showed that such filter banks could be used to build discrete orthogonal wavelet transforms [11].

While the constraint given by (9) must be satisfied if one is to create perfect reconstruction (PR) orthogonal filter banks and wavelets, it does not affect the creation to create non-orthogonal PR systems. In fact, if one wishes to construct a linear phase PR filter bank (or, equivalently, a symmetric wavelet), one must give generally give up orthogonality and equation (9). By doing so one can create instead a biorthogonal system such that

$$H_0(z)G_0(z) + H_0(-z)G_0(-z) = 1. \quad (11)$$

Two different filters must now be designed such that (11) is satisfied-- the remaining two filters are still determined by (5). If we define $P(z) = H_0(z)G_0(z)$, then (11) becomes

$$P(z) + P(-z) = 1 \quad (12)$$

which can be satisfied by a half-band or Nyquist filter. Thus, one need only design such a filter and then factor it to create the desired orthogonal or biorthogonal system [8].

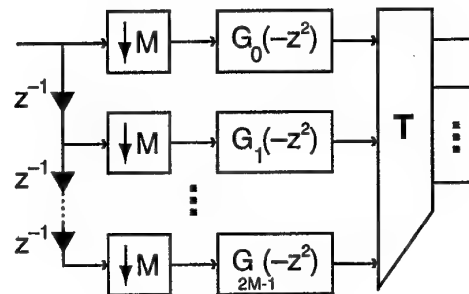


Figure 5: Polyphase implementation of cosine modulated M-band filter bank.

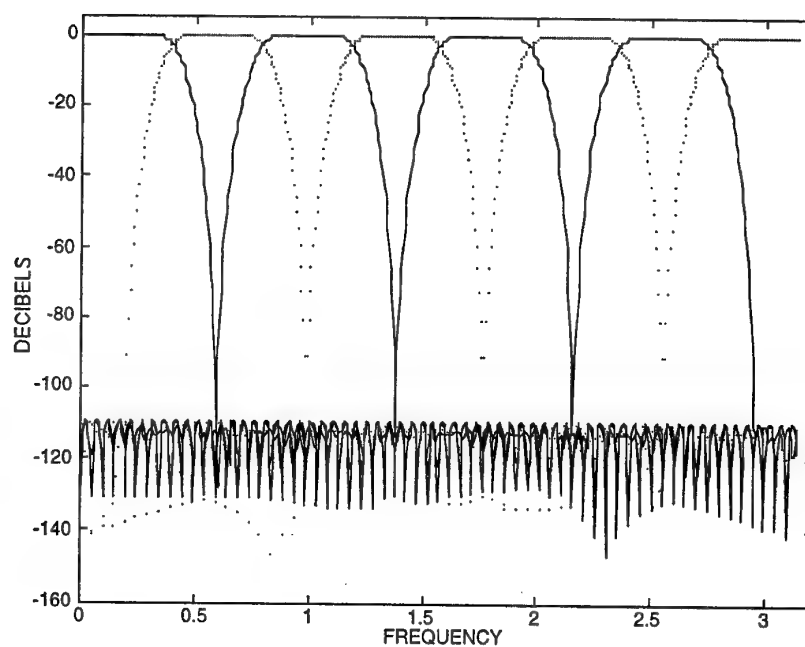


Figure 6: 8-band pseudo QMF bank. Note that only the transition bands of adjacent filters overlap.

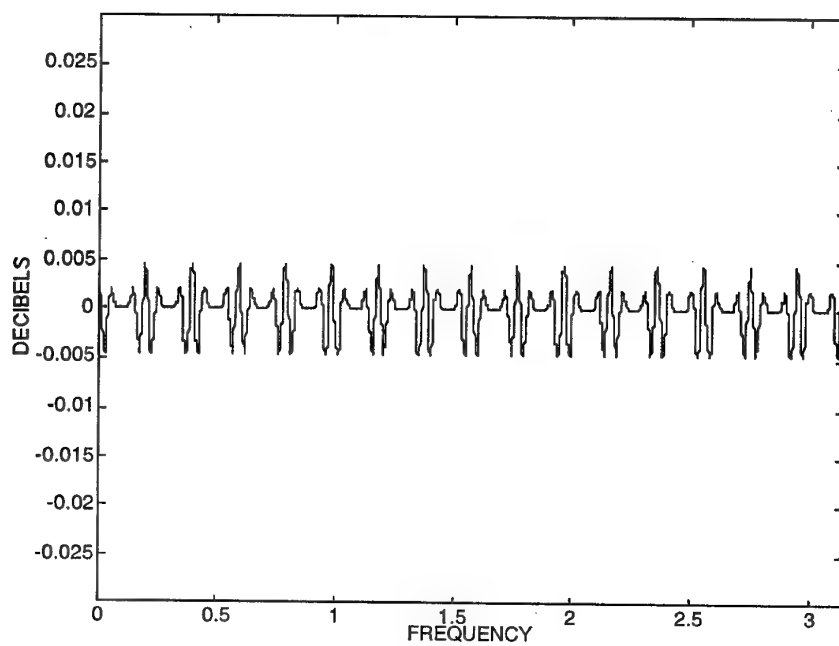


Figure 7: Frequency response of the 8-band analysis/synthesis system whose analysis bank is given by Fig. 6.

3.2 M-Band Systems

Often it is desirable to split a signal into more than 2 frequency bands. While one can implement an M-band decomposition by successively applying the 2-band filter bank discussed in the last section, it is often more efficient to implement it directly. In fact, the most efficient maximally decimated M-band filter bank capable of achieving good frequency discrimination is the cosine modulated filter bank [8]. Here, one designs a single lowpass prototype filter and then implements the remaining filters by using cosine waveforms to modulate this prototype to higher frequency bands. Figure 5 shows the computationally efficient polyphase implementation of an M-band cosine modulated filter bank (the powers of z^{-1} denote delay elements or shift registers). Note that the filter coefficients belonging to the lowpass prototype filter $H(z)$ are uniformly distributed amongst the polyphase subfilters $G_k(z^2)$. Thus, the complexity of implementing all M filters is equal to that of the original prototype filter $H(z)$ plus the cost of implementing the $M \times 2M$ transform T . The elements of the transform matrix T are given by

$$t_{kn} = 2 \cos\left(\frac{\pi}{M}\left(k + 0.5\right)\left(n - \frac{N}{2}\right) + \theta_k\right), \quad (13)$$

where $\theta_k = (-1)^k \pi/4$ and N is the order of $H(z)$. This transform can be implemented very efficiently using a $2M$ -point FFT plus a few additional operations.

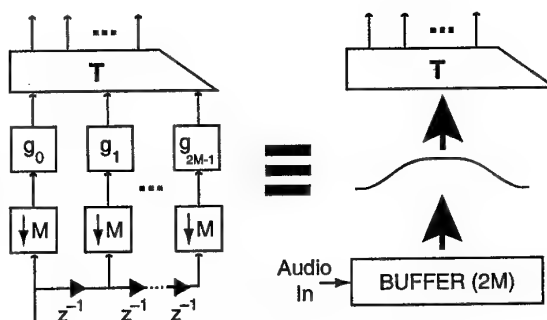


Figure 8: Relationship between LOT and modulated cosine filter bank.

The first filter banks designed along the lines of Fig. 5 were called pseudo-QMF or generalized QMF banks, and they did not achieve perfect reconstruction. Instead, they only cancel aliasing between adjacent frequency subbands, and they introduce a small amount of amplitude distortion into the reconstructed output. Figure 6 shows the analysis filters of an 8-band pseudo QMF bank. Note that only the transition bands of adjacent filters overlap-- the passbands and transition bands of all non-adjacent filters are in the stopband. Thus, it really is sufficient just to cancel aliasing between adjacent bands. Figure 7 shows the frequency response of the combined analysis/synthesis system. While there is some amplitude distortion, its magnitude is extremely small. Thus, despite its minor flaws, the pseudo-QMF bank has been widely used in a variety of applications, most notably MPEG (motion picture experts group) audio Layers 1, 2, and 3 (MP3). More recently, various authors have found ways of designing the prototype filter $H(z)$ so that perfect reconstruction can also be achieved within the framework of Fig. 5 [12].

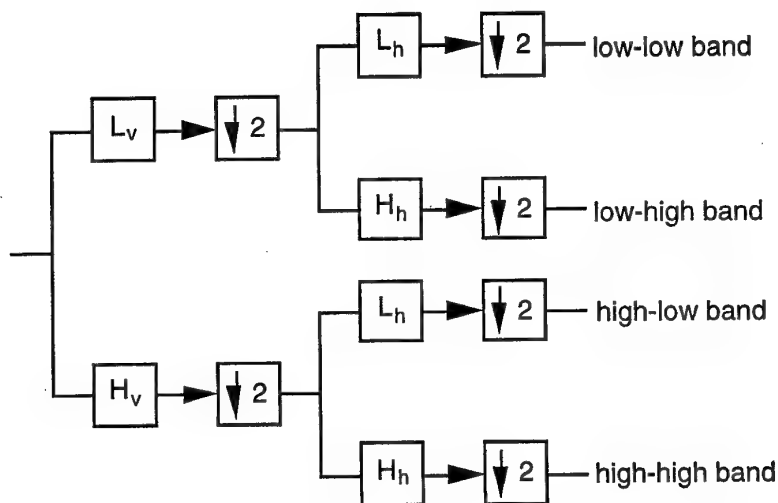


Figure 9: 4-band 2D multirate filter bank-- one level of wavelet decomposition.

It should also be noted that the cosine modulated filter bank of Fig. 5 can be specialized to the lapped orthogonal transform (LOT) or Local Cosine Transform (LCT). To see this, consider Fig. 8. Here, the polyphase subfilters have been replaced by single scalar coefficients. These coefficients are simply the window weights required to implement a LOT having a 50% overlap between adjacent windows. In the figure, the polyphase network acts exactly like a shift register in which half of the samples are replaced with new samples every clocking cycle.

4. Applications

4.1 Wavelet-Based Image Compression

The current standards for image and video compression, JPEG (joint photography experts group) and MPEG, are both transform coders centered around an 8x8 blocked DCT [1], [2]. More recently, image compression algorithms based on the wavelet transform have gained prominence in both research and development because of their ability to operate effectively over a wide range of compression ratios [13], [14]. In fact, a single wavelet-based algorithm can be constructed which operates efficiently for both lossless and highly lossy compression [15], [16]. The basic 2-dimensional (2D) wavelet transform used for image compression is constructed by cascading two 1D filter banks of the type shown in Fig. 4 so that the first one operates in the vertical direction and the second in the horizontal direction. The 4-band 2D filter bank thus created is shown in Fig. 9 where the subscripts denote the direction of operation of the low (L) and high (H) pass filters. To create a complete wavelet transform, we iterate the filter bank shown in Fig. 9, successively decomposing the low-low band. Figure 10 shows the subband or wavelet coefficient mapping which results from 3 such iterations. To further illustrate this point, Figure 11 shows the wavelet decomposition of an actual image.

The process of wavelet-based compression is illustrated in generic terms by Fig. 12. First the image is transformed and then the transform coefficients are appropriately quantized (eliminating information content) and coded (eliminating redundancy and converting into bits). To reconstruct an approximation of the image, we decode the bit stream and perform an inverse wavelet transform (IDWT) which is simply the synthesis filter bank corresponding to the analysis bank of Fig. 9.

Of particular interest is a class of wavelet-based image compression algorithms called 'embedded' coders [13]-[16]. In an embedded coder, a bit stream is generated and transmitted in order of importance-- i.e., the most important bits are sent first. Figure 13 illustrates this concept. Embedded bit streams are useful in a number of ways: 1) if transmission terminates prematurely, we can still reconstruct an image of the best possible quality with the bits we received, 2) unequal error protection coding is easily achieved in a dynamic fashion, and 3) communications channels with fixed bit rates are easily supported. All of these traits are useful in military applications because the communications channels of interest are primarily RF and are likely to be operated in an unstable and noisy environment.

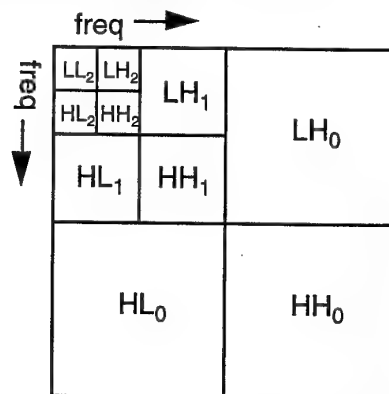


Figure 10: Wavelet coefficient or subband mapping.



Figure 11: Wavelet decomposed image.

While a number of embedded coding algorithms have been developed [13]-[15], we describe here the embedded zerotree wavelet (EZW) approach which was the first developed and is still one of the best [13]. The fundamental observation around which this coding algorithm is centered is that there is a strong correlation between insignificant coefficients at the same spatial locations in different wavelet scales-- i.e., if a wavelet coefficient at a coarser scale is zero, then it is more likely that the corresponding wavelet coefficients at finer scales will also be zero. Figure 14 shows a 3-level, 2D wavelet decomposition and the links which define a single zerotree structure. If the wavelet coefficient at a given scale is zero along with all of its descendants (as shown in Fig. 14), then a special symbol indicating a zerotree root (ZTR) is transmitted, eliminating the need to transmit the values of the descendants. Thus, the correlation of insignificance across scales results in a net decrease in the number of bits transmitted.

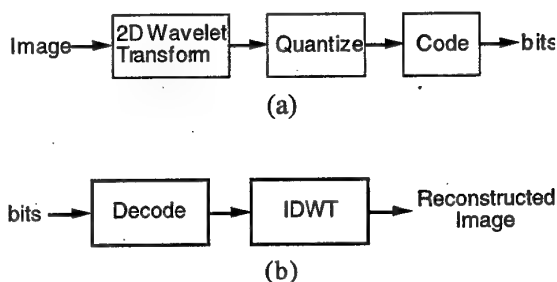


Figure 12: Wavelet compression and decompression.

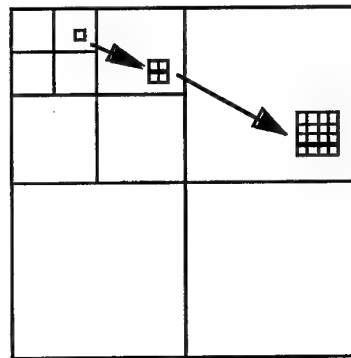


Figure 14: Parent-child relationships amongst zerotree coefficients.

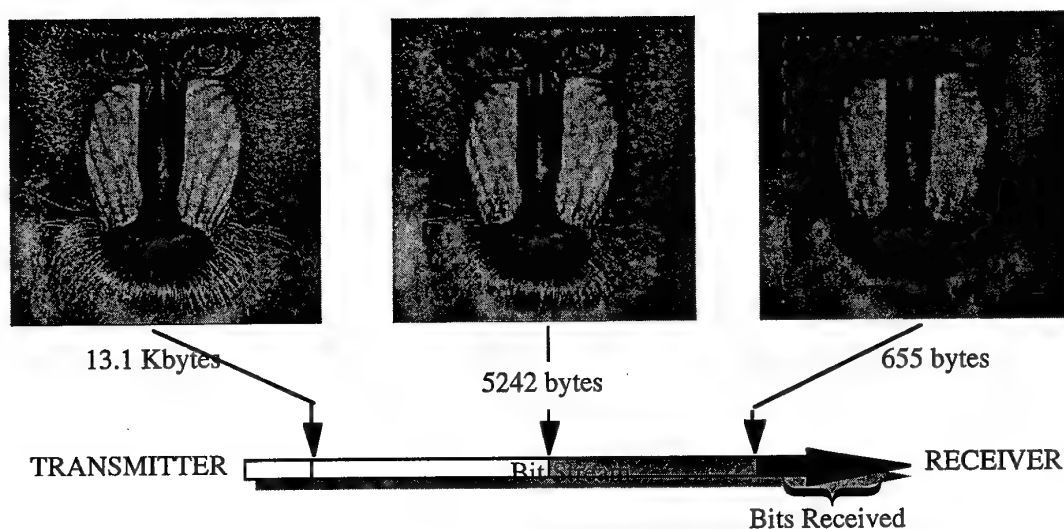


Figure 13: Embedded image compression.

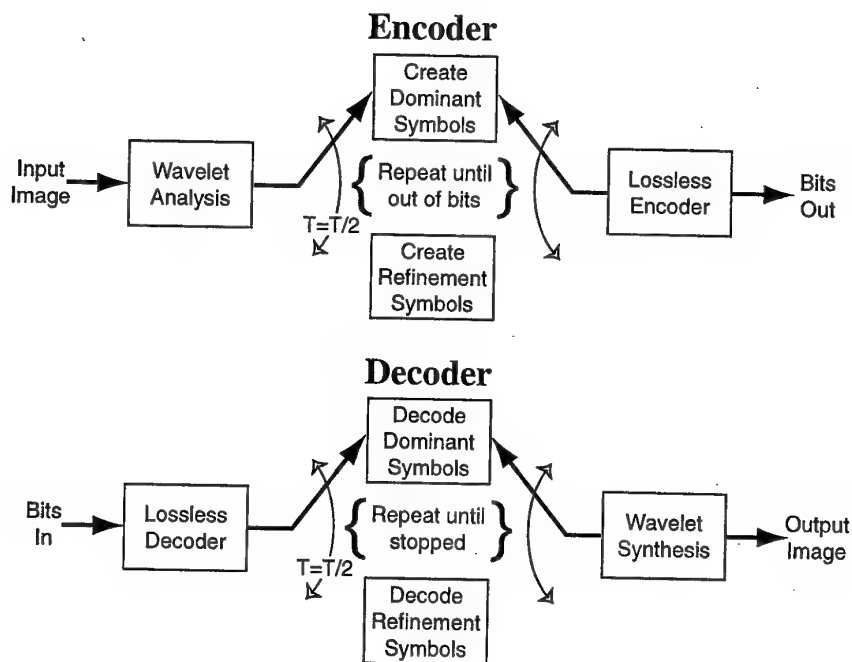


Figure 15: Encoder and decoder for an embedded image compression algorithm.

In order to generate an embedded code (where information is transmitted in order of importance), Shapiro's EZW algorithm scans the wavelet coefficients in a bit-plane fashion. Starting with a threshold determined from the magnitude of the largest coefficient, the algorithm sweeps through the coefficients, transmitting the sign (+ or -) if a coefficient's magnitude is greater than the threshold (i.e., it is significant), a ZTR if it is less than the threshold but the root of a zerotree at the coarsest possible scale, or a 0 otherwise--this is the dominant pass. Next, for the subordinate pass all coefficients deemed significant in the dominant pass are added to a second subordinate list which is itself scanned, adding one bit of resolution to the decoder's representation of each significant coefficient. Symbols generated by these two passes are then passed through a lossless arithmetic coder to extract further statistical redundancy. After this, the threshold is halved and the two passes are repeated with those coefficients having been found significant previously being replaced by zeros in the dominant pass (so that they do not inhibit the formation of zerotrees in subsequent iterations). The process continues until the bit budget is exhausted; at this point, the encoder transmits a stop symbol and its operation is terminated.

The decoder, on the other hand, simply accepts the bit stream coming from the encoder, progressively building up the significance map and

the subordinate list in the exact same way as they were created by the encoder. Because of this precise synchronization, the resolution enhancement bits transmitted during the subordinate pass do not need any location specifiers-- the decoder knows the exact transmission order of these bits because it has reconstructed the same subordinate list as the encoder had at that point in the process. The encoding and decoding processes described here are illustrated by the block diagrams shown in Fig. 15. Figure 16 illustrates the advantages of wavelet-based compression over JPEG at high compression ratios.

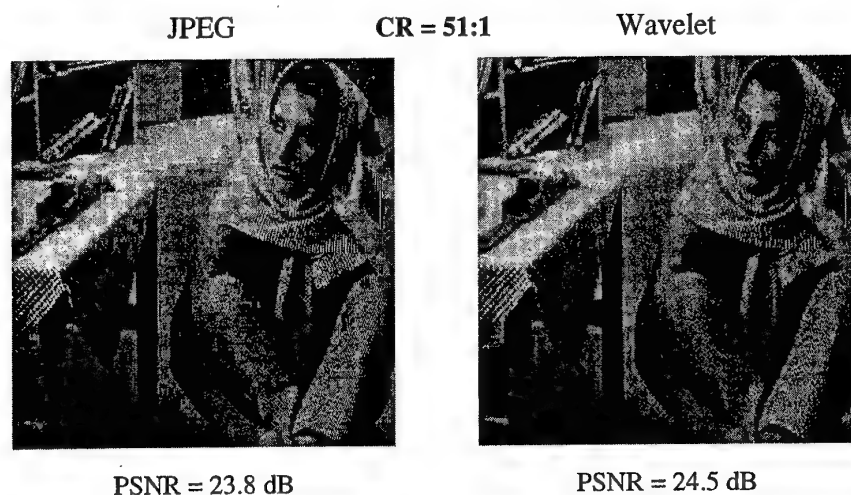


Figure 16: JPEG versus EZW at a compression ratio of 51:1. Objective comparison metric is peak signal to noise ratio (PSNR).

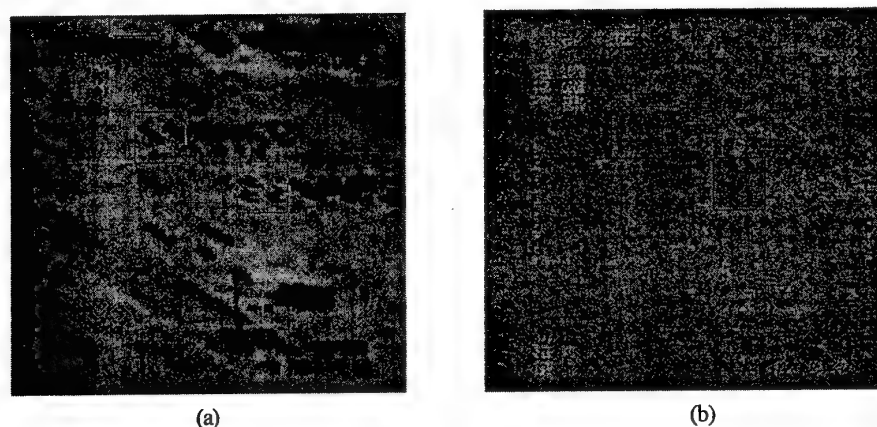


Figure 17: (a) Reconstruction of image compressed by 160:1 ratio. Again, squares have been added to highlight enhanced regions. (b) Error residual between reconstructed and original images where white areas in residual denote large errors.

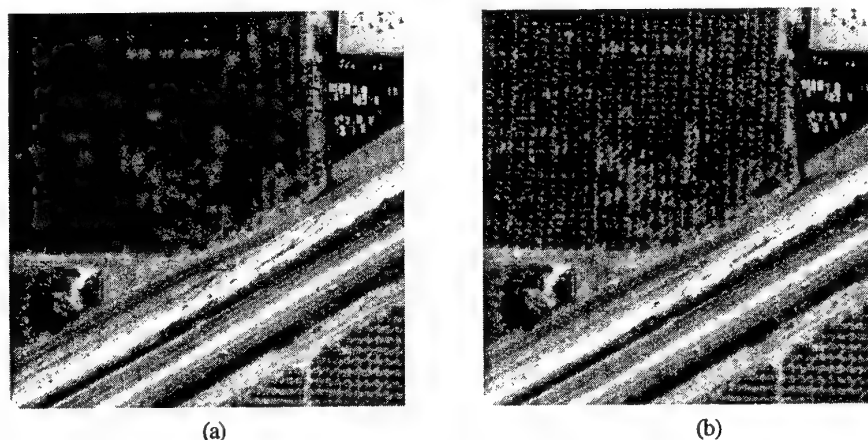


Figure 18: (a) Compressed with space-frequency weighting by 32:1 ratio; (b) Compressed without weighting. Note that our texture detection algorithm is designed to be selective-- it only identified the orchard in the upper left-hand corner of the image but not the one in the lower right-hand corner.

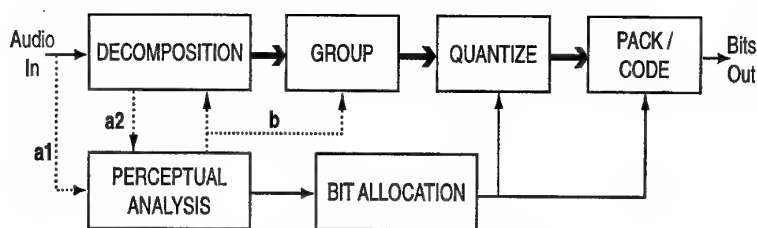


Figure 19: Generic audio encoder.

One should note that the use of wavelet transforms for image (or video) compression does not preclude spatial significance weighting. In other words, if you know that certain regions of the image contain more (or less) important content, you can have the encoder allocate more (or less) resolution to those areas. In fact, within the framework of the wavelet-decomposed image, the encoder can actually control the allocation of resolution in both space *and* frequency-- within certain constraints, of course. In [17] we combined an embedded image compression algorithm with a maximum average correlation height (MACH) feature detector to allocate resolution within the image frame. Figure 17 shows the result of this feature-based compression algorithm when four different resolution weightings are available. Similarly, we can also decrease the resolution of areas that our coder determines to be of low interest (in order to make more bits available for the rest of the image). In Fig. 18, we have forced the encoder to discard higher frequency wavelet coefficients within the orchard on the assumption that the details of the orchard are not of great interest. Although not obvious in the figure, the areas outside the orchard are represented with higher fidelity in Fig. 18b than 18a.

4.2 Wideband Audio

Another area in which multirate filter banks have had great success is wideband audio compression. Because of the high dynamic range of typical wideband audio signals (e.g., classical music), a highly adaptive decomposition and bit allocation framework (dynamically adapted to the signal power levels at a given time and within a given frequency band) is required. The basic time-frequency decomposition used in virtually every wideband audio coder generated is the cosine modulated filter bank in either the general polyphase form of Fig. 5 or the LOT form of Fig. 6. To allocate bits to the different frequency bands at a given time, the algorithm uses knowledge about perceptual masking in the human ear [18].

Figure 19 shows the block diagram of a generic subband (transform) coding algorithm which can adapt its quantization (and possibly its decomposition) to optimize the perceived quality

of its reconstructed audio. The dotted lines represent data exchanges which do not occur in all implementations. For example, all three of the MPEG 1 audio coders (Layers 1, 2, and 3) use a separate FFT to perform the frequency analysis required to do the bit allocation (indicating that path *a1* is active) while Dolby AC-3, Philip's digital compact cassette (DCC), and Sony's MiniDisc use only the outputs of their signal decompositions (path *a2*). The 'Perceptual Analysis' block computes the masking estimates which are required by the 'Bit Allocation' block to ensure that quantization errors in the reconstructed audio are inaudible. Based on this analysis of the signal, some audio coders also have the ability to alter their decompositions and the corresponding coefficient groupings (path *b*) to prevent the introduction of pre-echoes into the decoded audio. Note that the decoder simply inverts the operations of the encoder block by block to reconstruct an approximation of the input audio.

Perceptually transparent coding is accomplished primarily by exploiting the various masking properties of the human ear, specifically: the absolute threshold of hearing, simultaneous frequency masking, forward (temporal) masking, and backward masking. First, any frequency component of the signal whose power falls below the absolute threshold of human hearing need not be transmitted. This threshold is lowest between 2 and 4 kHz and goes up rapidly above 15 kHz. Next, if a small amplitude tonal signal occurs at the same time as a larger one of similar frequency, the smaller signal will be masked. This is called simultaneous masking and is specified in terms of critical bands which are defined on the bark scale [18]. These critical bands define the frequency resolution of the human auditory system-- from 0 to 500 Hz there are 5 uniform critical bands while above 500 Hz the width of each band expands by approximately 1/3 per octave. The effectiveness of the masking decreases by about 8 dB/bark for critical bands above the masker and 25 dB/bark for those below it, and it also depends strongly on the tonality of the input since pure tones mask each other much more effectively than noise-like signals. To estimate tonality, the Spectral Flatness Measure (SFM)-- basically the logarithm of the power spectrum's

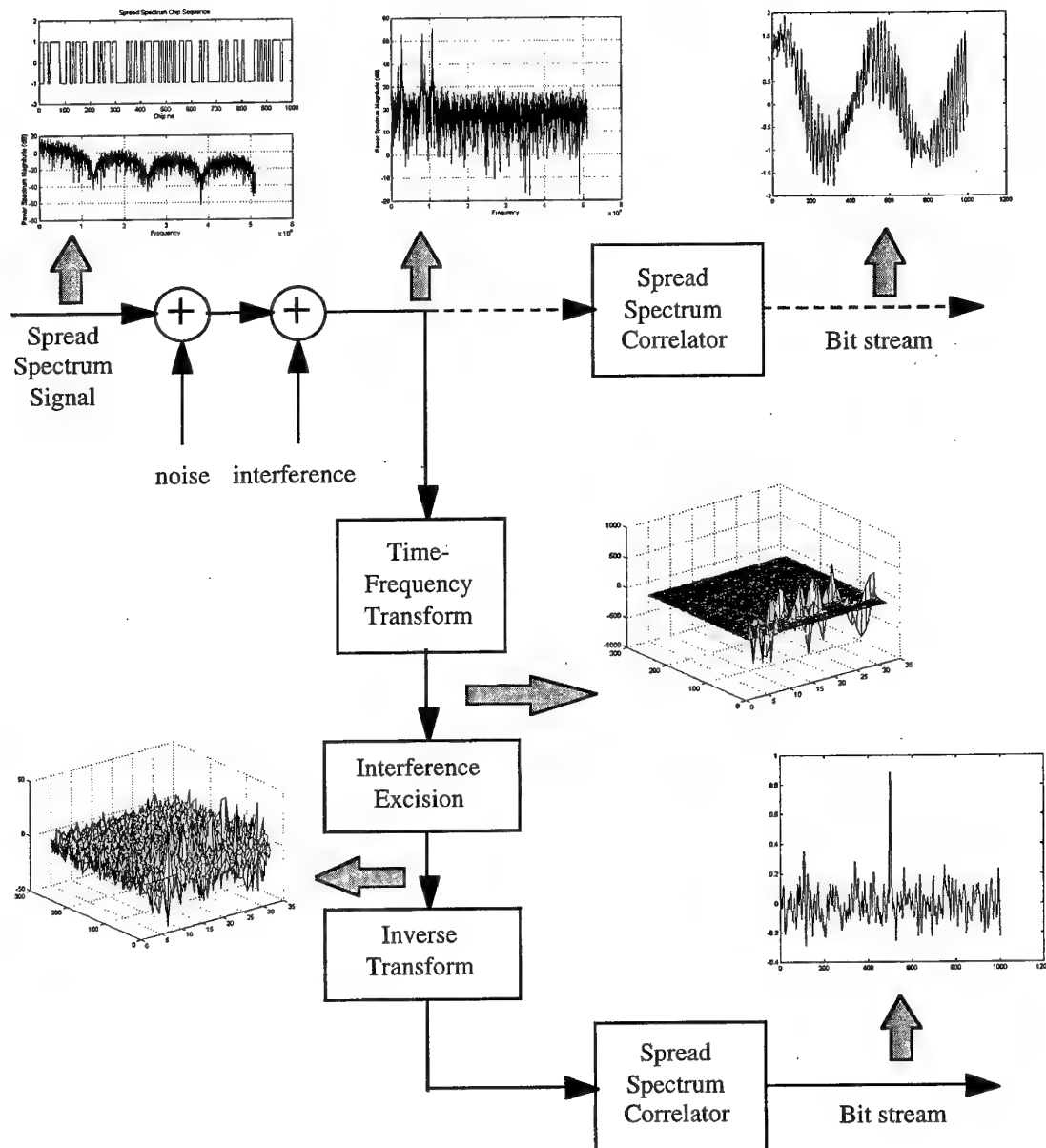


Figure 20: Time-frequency based interference excision. Original signal enters noisy communications channel at left. Signal output (to the right) at the top is the bit stream produced if interference is not excised while the one at the bottom has been denoised. Note the clear correlation spike in the bottom signal (indicating a good signal lock) and the lack of such a spike in the upper signal.

geometric mean divided by its arithmetic mean-- is generally employed [19]. Specifically, a ratio of the current SFM to the SFM of a maximally tonal input is used to compute the tonality of the current block of samples, and this tonality coefficient biases the masking threshold upward for highly tonal signals or downward for noise-like signals.

The final perceptual effects which must be considered in the design of the coding algorithm are forward and backward temporal masking. Forward masking occurs when the masking signal ends before the masked signal begins while backward masking is the exact opposite. Perceptual studies have shown that forward masking is the more effective of the two by a wide margin [18]. While most of the currently available coding algorithms claim to 'exploit' forward and backward masking, this statement is somewhat misleading. Explicitly, they exploit *simultaneous* masking to achieve bit rate reductions through adaptive bit allocation while implicitly exploiting forward masking to conceal the effects of time-frequency blocking on the quantized coefficients. In other words, if the masking signal contained within the block of coefficients ends prematurely, the quantization noise will still be concealed. The situation with backward masking, however, is entirely different since this phenomenon is highly localized around the leading edge of the masker. If blocks of coefficients representing a fixed time-frequency subdivision of the signal are jointly coded, then it is possible for pre-echo to be introduced into the reconstructed audio by the occurrence of a large masker in latter parts of a block. Thus, the goal of the coding algorithm is not so much to exploit backward masking as to compensate for its limitations. In fact, the entire motivation for using temporally adaptive transformations in the encoder (path **b** in Fig. 19) comes from the need for increased time localization of the quantization errors during sharp attacks (i.e., sudden increases in the short time power spectrum of the audio input).

4.3 Time-Frequency Interference Excision

The same transforms we applied in the last section to the rather non-military application of wideband audio compression can also be applied to the problem of removing narrow band interference from a wideband signal [20]. To do this, we first transform the signal using the LOT of Fig. 6 and then analyze the frequency subband, looking for any large concentrations of energy. Since our desired signal is wideband, its energy will not be particularly concentrated in any given band. The energy from a narrowband interference, on the other hand, *will* be concentrated into certain bands, even if its frequency is hopping or chirping (within certain bounds). Once a band containing a potential interference source is detected, we alter its transform coefficients so as to suppress the interference without introducing new types of distortions into the signal. Note that this is more complicated than simply zeroing the offending coefficients since these zeros themselves can introduce false information into the output signal. Finally, the transform coefficients-- possibly modified-- are inverted, and the denoised signal is then sent on for detection or further processing. A block diagram of the complete system is shown in Fig. 20.

Superimposed on the excision system in Fig. 20 is an example illustrating the advantages of excision. The input spread spectrum signal enters from the left and is corrupted by noise and narrowband interference (effects of the communications channel). The upper branch outputting the signal to the right simply correlates the corrupted signal with the original without denoising. Note that the correlation peak is totally obscured by noise. The signal coming out of the lower correlator, on the other hand, has had its interference detected and removed using time-frequency methods [21]. Here, the correlation peak at the output is highly pronounced, indicating that we can easily decode the information bits modulated onto this waveform.

One can also use a windowed Fourier transform in this application. The disadvantage of doing this, however, is summarized by the Balian-

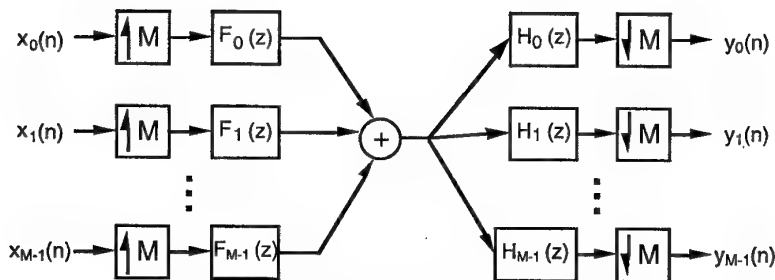


Figure 21: COFDM transmission system.

Low Theorem: it is impossible to design a windowed Fourier transform which simultaneously achieves good time and frequency localization but does not expand the sampling rate of the signal in the transform domain. Thus, to prevent the excision process from introducing artifacts into the reconstructed signal, one must generate and process approximately 50% more transform domain coefficients using a Fourier transform than with the lapped orthogonal cosine transform.

4.4 Code Orthogonal Frequency Division Multiplexing (COFDM)

The basic idea of COFDM is to combine a large number of low bandwidth information sources into a single wide bandwidth signal for transmission [6]. This can be accomplished using the dual structure to the PR filter bank-- the perfect transmultiplex. Such a system is shown in Fig. 21 where M narrow bandwidth signals are combined into one signal having M -times more bandwidth. Note that while each of the M input signals is mostly confined to a specific frequency band within the combined signal, there is some overlap because the digital filter are not ideal. Despite this, if the filters $F_k(z)$ and $H_k(z)$ are designed correctly, all cross-talk will be cancelled and no amplitude or phase distortion will be introduced into the output signals: i.e., $y_k(n) = x_k(n)$ for all k . It is easily shown that if filters $F_k(z)$ are the synthesis filters for an M -band PR filter bank and $H_k(z)$ are the corresponding analysis filters, then the outputs of Fig. 21 will exactly equal its inputs, resulting in a perfect transmultiplexer.

If the impulse responses of the set of filters $\{f_k(n)\}$ are all mutually orthogonal to each other, then the system illustrated by Fig. 21 is truly COFDM. Note that orthogonality of the synthesis filter set $\{f_k(n)\}$ also guarantees orthogonality of the analysis filter set $\{h_k(n)\}$. If a PR cosine modulated filter bank of the type shown in Fig. 5 is used here, then the filter set $\{f_k(n)\}$ is indeed orthogonal. Higher order filters results in better frequency confinement of the original signal

$x_k(n)$ within the composite signal, but they also increase the complexity of the system. To make M as large as possible for a given implementation complexity, the LOT of Fig. 6 is often used. This results in relatively poor frequency localization of a given narrowband input within the broadband composite signal but does not appear to cause any serious problems within the context of this application.

Europe has already selected this modulation scheme for Digital radio broadcasting and will likely select it for digital television as well. Why? A major advantage is that it is much more robust to multipath phenomena and other forms of frequency dependent interference. Specifically, one can adjust the information flow rate going into each of the M input channels so that every channel operates at exactly its maximum capacity. Thus, a channel that is subject to a great deal of interference would have much less data capacity than one which is not. The data rate for each channel can be adjusted by altering the amount of error correction used in it-- i.e., the low rate, interference prone channel expends most of its raw capacity on error protection while the high rate, interference free channel uses more of its capacity to transmit actual data bits. In some cases, there might also be a feedback path from the receiver to the transmitter, allowing the system to adapt to the changing RF environment. Note that numerous variations on this concept have been proposed including wavelet modulation and discrete tone modulations. Both of these use the same basic concept as COFDM, but with different modulation kernels [22].

4.5 Time-Frequency Scrambling

It has been shown in [5] how multirate filter banks can be used very effectively to implement voice scrambling systems based on time-frequency permutation. The most general form of such a system is shown in Fig. 22 where $P(z)$ is the permutter (it is a function of z because it has, in general, memory). Since the subband coefficients are only reordered by $P(z)$ and not, in the-

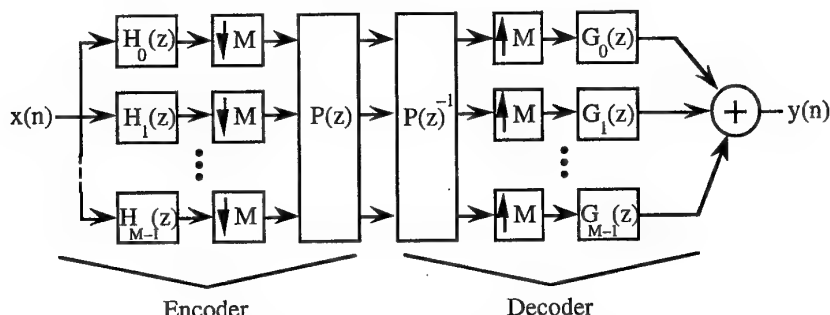


Figure 22: Time-frequency scrambling.

ory, quantized, any perfect reconstruction filter bank can be used for the analysis and synthesis operations in Fig. 22 without loss in the quality of the reconstructed signal. If the signal input to the encoder, $x(n)$, has a frequency distribution which is known to the pirate (e.g., mostly lowpass), then the pirate will be able to use this knowledge to more easily break the permuter's scrambling algorithm. This liability can be eliminated at considerably higher cost by passing the permuted signal from $P(z)$ through the synthesis bank and transmitting the output. In this case, the decoder must first pass its received signal through an analysis bank, process it with the inverse permuter, $P(z)^{-1}$, and, finally, reconstruct the unscrambled signal. If the filter banks used in the process are perfect reconstruction, the unscrambled signal will still match the original at the end of this process (barring multiplication roundoff and coefficient quantization errors).

Time-frequency scrambling using maximally decimated multirate filter banks has a number of advantageous properties. Obviously, it obscures the signal in both time and frequency, making it very difficult to decode it even if your opponent has a priori knowledge about the statistics of the signal (at least as long as the filters and synchronization intervals are not known exactly). Furthermore, since the filter bank is maximally decimated, the scrambler does not increase the bit rate of the signal. On the minus side, the delay introduced by the permuter $P(z)$ for any given sample must be limited to some maximum value for real-time applications. With 2-way voice communications, for example, it becomes impossible to carry on a conversation if the throughput delay or latency grows too large. If little system latency is allowed, the delay constraint on $P(z)$ permits an opponent to limit his key search and thus decode the data more quickly. Time-frequency scrambling systems are most useful for protecting data whose importance is highly time dependent—i.e., data which is immediately useful but has little long term value. For other kinds of sensitive data, such time-frequency systems probably do not provide adequate protection.

5. Conclusions

We have discussed some of the basic concepts behind digital multirate systems, focusing on the area of maximally decimated filter banks and wavelets. In particular, we have analyzed the 2-band PR filter bank which implements orthogonal and biorthogonal wavelets. We have also considered the important cosine modulated filter bank, a special case of which is the lapped orthogonal transform or local cosine transform. Finally, we have summarized a number of communications-related applications for such filter banks including wavelet-based image compression, wideband audio compression, noise excision, modulation,

and scrambling. Of course, many other applications for such multirate systems also exist including adaptive filtering, pattern recognition, and channel equalization. In short, multirate digital signal processing plays an important role in modern communications systems, and it will likely become even more critical as digital connectivity increases in the future.

Acknowledgement

The author would like to thank Dr. Grant Hanson for contributing material on time-frequency noise excision and the sponsors at the Office of Naval Research who have facilitated this research.

References

- [1] G.K. Walley, "Overview of the JPEG (ISO/CHIT) still image compression standard," *Proc. of the SPIE*, Vol. 1244, pp. 220-33, 1990.
- [2] D.J. LeGall, "The MPEG video compression algorithm: a review," *Proc. of the SPIE*, Vol. 1452, pp. 444-57, 1991.
- [3] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice Hall PTR, Englewood Cliffs, NJ, 1995.
- [4] P. Noll, "MPEG digital audio coding," *IEEE Signal Processing Magazine*, pp. 59-81, Sept. 1997.
- [5] R.V. Cox, D.E. Bock, K.B. Bauer, J.D. Johnston, and J.H. Synder, "The analog voice privacy system," *AT&T Technical Journal*, Vol. 66, No. 1, Jan.-Feb. 1987, pp. 119-131.
- [6] W.Y. Zou and Y. Wu, "COFDM: an overview," *IEEE Trans. on Broadcasting*, Vol. 41, No. 1, pp. 1-8, March 1995.
- [7] A.V. Oppenheim and R.W. Schaffer, *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1989.
- [8] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, Englewood Cliffs, NJ, 1993.
- [9] C.D. Creusere and G. Hewer, "Wavelet-based nearest neighbor pattern classification using scale sequential matching," *Proc. Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Nov. 1994, pp. 1123-1127.
- [10] M.J.T. Smith and T.P. Barnwell, "Exact reconstruction techniques for tree-structured subband coders," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-34, No. 3, June 1986, pp. 434-440.
- [11] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Comm. on Pure*

and *Applied Math.*, Vol. XLI, 1988, pp. 909-996.

[12] H.S. Malvar, "Modulated QMF banks with perfect reconstruction," *Electronics Letters*, vol. 26, pp. 906-7, June 1990.

[13] J.M. Shapiro, "Embedded zerotree image coding using zerotrees of wavelet coefficients," *IEEE Trans. on Signal Processing*, Vol. 41, No. 12, Dec. 1993, pp. 3445-3462.

[14] A. Said and W.A. Pearlman, "An image multiresolutional representation for lossless and lossy compression," *IEEE Trans. on Image Proc.*, Vol. 5, No. 9, Sept. 1996, pp. 1303-1310.

[15] A. Zandi, J.D. Allen, E.L. Schwartz, and M. Boliek, "CREW: Compression with reversible embedded wavelets," *Proc. Data Compression Conference*, 1995, pp. 212-221.

[16] C.D. Creusere, "Spatially partitioned lossless image compression in an embedded framework," *Conf. Rec. 31st Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1997, pp. 1455-59.

[17] C.D. Creusere and A. Van Nevel, "Automatic target recognition directed image compression," to appear in the *Journal of Aircraft*.

[18] E. Zwicker and H. Fastl, *Psychoacoustics*. Berlin: Springer-Verlag, 1990, pp. 141-147.

[19] J.D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Select. Areas Commun.*, vol. 6, pp. 314-23, Feb. 1988.

[20] J.H. Young and J.S. Lehnert, "Performance metrics for windows used in real-time DFT-based multiple-tone frequency excision," *IEEE Trans. on Signal Processing*, vol. 47, No. 3, pp. 800-12, March 1999.

[21] G. Hanson, "Adaptive time-frequency jammer excision using local cosine transforms," *Proc. 1999 National Technical Meeting & 19th Biennial Guidance Test Symposium* (classified volume), Jan. 1999, San Diego, CA.

[22] A.N. Akansu and X. Lin, "A comparative evaluation of DMT (OFDM) and DWMT (DSBMT) based DSL communications systems for single and multitone interference," *Proc. Int. Conf. on Acoustics, Speech, and Signal Proc.*, Vol. 6, pp. 3269-72, 1998.

Multisensor Signal Processing in the Framework of the Theory of Evidence

Alain Appriou

ONERA

BP 72, 92322 Châtillon Cedex, France
Tel. : +33 1 46 73 49 19 ; Fax : +33 1 46 73 41 67
E-mail : appriou@onera.fr

Summary: In most of the requirements met in situation assessment, multisensor analysis has to be able to recognize in adverse conditions one situation out of a set of possibilities concerning for instance either localization, identity, or matching hypotheses. To do so, it uses measurements of more or less doubtful origin and prior knowledge that is understood to be often poorly defined, and whose validity is moreover difficult to evaluate under real observation conditions. The present synthesis proposes a generic modeling of this type of information in the framework of the theory of evidence, with closer attention being paid to the different natures of data processed in common cases. This modeling is then used to elaborate processing methods able to face specific problems that may arise when multisensor systems are implemented to achieve functions like detection, classification, matching of ambiguous observations, or tracking. Crucial practical problems are more specifically dealt with, such as appropriate combination processing and decision making, management of heterogeneous frames of discernment, and integration of contextual knowledge. Furthermore, the interest of a global functional approach at low level, possible in that framework, is emphasized.

1 Data Uncertainty in Multisensor Systems

Sensors are mainly associated in order to get benefit of their complementarity. Different kinds of advantages may be expected :

- ability to face a more important set of situations, as one sensor may be efficient while an other one is not because of particular counter-measures, physical phenomena, conditions of observation, or lack of suitable knowledge (learning,...) ;
- saving of time thanks to task sharing and cooperation between specific functions ;
- discrimination capacity improvement as a result of observation conjunction when only partial information is locally available (classification, localization,...).

Consequently, when analyzing a situation, the available sensors are most often used under conditions that induce uncertainties at different levels :

- measurements may be imprecise, erroneous, incomplete, or ill-suited to the problem,
- observations may be ambiguous, either in space or in time (e.g. position, velocity, or feature measurements provided by two different sensors are not necessarily related to a same object),
- prior knowledge (generated by learning, models, descriptions, and so forth) may be incomplete, poorly defined, and especially more or less representative of reality, in particular in light of the varying context.

Moreover, the disparity of the data delivered by the various sensors, which is intended to remedy the individual insufficiencies of each, requires a detailed evaluation of each of them, based on any exogenous information that might characterize their pertinence to the problem at hand and the context investigated, while such information is itself often very subjective and imprecise.

Theories of uncertainty offer an attractive federative framework in this context. But they run up against a certain number of difficulties in practice : interpretation and modeling of the available information in appropriate theoretical frameworks, choice of an association architecture and combination rules, decision principles to be adopted, constraints concerning the speed and volume of the necessary computations.

To provide solutions to these questions, we will first consider a generic problem in which we attempt to characterize the likelihood of I hypotheses H_i theoretically listed in an exhaustive and exclusive set E . These hypotheses may typically concern the presence of entities, target or navigation landmark identities, vector or target localization, or the status of a system or of a situation.

Such a likelihood function may then be integrated either into :

- a choice strategy, to declare the most likely hypothesis (target identification, intelligence, and so on),
- a filtering process (such as target tracking or navigation updating),
- a decision aid process for implementing means of analysis, electronic warfare, or intervention.

The likelihood functions we want have to be developed from the data provided by J sensors S_j . Each of them is assumed to be associated with processes that extract a measurement or a set of measurements s_j , pertinent to the targeted discrimination function, from the raw signals or images it generates.

The developments presented are conducted in the theory of evidence framework [1], which provides the broadest and best-suited tools for the interpretation and the processing of the data considered. It is also the most federative in terms of synergy between the different theoretical frameworks that may be involved together. Appendix A gives a few basic notions that will be used in the following.

We first propose a common solution to the generic modeling problem introduced formerly, that is afterwards particularized when closer attention is paid to the different natures of data processed in common cases. This modeling is then used to elaborate suitable classification methods, thanks to appropriate combination processing and decision making. Furthermore, specific problems met in multisensor systems are considered, such as management of heterogeneous frames of discernment, integration of contextual knowledge, and matching of ambiguous observations. Finally, the interest of a global functional approach at low level, possible in that framework, is emphasized with the implementation a tracking process that manages directly discrimination features.

2 Modeling of Input Data

2.1 Generic Model

In the framework of the generic problem we will be considering first, we assume that each measurement s_j can be used to generate I criteria C_{ij} , on the basis of any *a priori* knowledge, having values in $[0,1]$ capable of characterizing the likelihood of each hypothesis H_i . A quality factor q_{ij} with values in $[0,1]$ is also associated with each likelihood C_{ij} . Its purpose is to express the aptitude of the criterion C_{ij} to discriminate the hypothesis H_i under the given observation conditions, on the basis of a dedicated learning process or exogenous knowledge. This factor includes mainly the confidence that can be accorded to the validity of the *a priori* knowledge used for generating C_{ij} . As concerns, for example, the representativity of a learning process in a varying context, it will typically depend on the quality, volume, and suitability of the available preliminary data as regards the situation effectively met.

Furthermore, we consider here the practical case of interest when the criteria C_{ij} are generated by separate information channels, in agreement with the fact that they are characterized by different levels of reliability q_{ij} . We also assume that we are in the most frequently

encountered context where the criteria C_{ij} taken separately are always at least of refutation value, in the sense that, when zero, this guarantees that the associated hypothesis H_i is not verified.

This leads to a formal construction of the problem on the basis of two axioms [2], [3] :

Axiom 1 : Each of the $I*J$ pairs $[C_{ij}, q_{ij}]$ constitutes a distinct source of information having the focal elements H_i , $\neg H_i$, and E , in which the frame of discernment E represents the set of the I hypotheses H_i .

Axiom 2 : When $C_{ij} = 0$, C_{ij} being valid ($q_{ij} = 1$), we can assert that H_i is not verified.

Axiom 1 requires that $I*J$ mass functions $m_{ij}(\cdot)$ be generated from the $I*J$ respective pairs $[C_{ij}, q_{ij}]$. For each, the mass of focal elements H_i , $\neg H_i$, and E is at first defined by the value of the corresponding criterion C_{ij} , which can be interpreted only in terms of credibility or plausibility of H_i . Axiom 2 then limits the number of allowable interpretations to two. The first interpretation leads to :

$$Cr_{ij}(H_i) = 0 \quad \text{and} \quad Pl_{ij}(H_i) = C_{ij} \quad (2.1)$$

and the second to :

$$Cr_{ij}(H_i) = Pl_{ij}(H_i) = C_{ij} \quad (2.2)$$

Then, including the confidence factor q_{ij} for C_{ij} by discounting at the rate $(1-q_{ij})$ provides the desired mass function $m_{ij}(\cdot)$. This leads to the two possible models :

Model 1 :

$$m_{ij}(H_i) = 0 \quad (2.3)$$

$$m_{ij}(\neg H_i) = q_{ij} * (1 - C_{ij}) \quad (2.4)$$

$$m_{ij}(E) = 1 - q_{ij} * (1 - C_{ij}) \quad (2.5)$$

Model 2 :

$$m_{ij}(H_i) = q_{ij} * C_{ij} \quad (2.6)$$

$$m_{ij}(\neg H_i) = q_{ij} * (1 - C_{ij}) \quad (2.7)$$

$$m_{ij}(E) = 1 - q_{ij} \quad (2.8)$$

A mass function $m(\cdot)$ synthesizing all the evaluations is then obtained by computing the orthogonal sum of the different mass functions $m_{ij}(\cdot)$ in the framework of each model :

$$m(\cdot) = \bigoplus_{i,j} m_{ij}(\cdot) \quad (2.9)$$

It should be noted that Model 1 is consonant. Furthermore it satisfies the minimum of specificity measure [4].

The practical determination of the C_{ij} and q_{ij} terms is of course a problem specific to the type of application at hand. The following in this section 2 provides expressions of C_{ij} for the different natures of data processed in common situations. The determination of q_{ij} is discussed in sections 3 and 5.

2.2 Model With Statistical Learning

We have to consider different kinds of relation between the sensor observations and the discrimination features that are characterized by previous learning.

2.2.1 Precise and Reliable Observation

The problem dealt with here assumes that each of the measurements s_j is directly one of the discrimination features exploited or a deterministic function of it, so that for each of them a learning of their *a priori* probability distribution $p(s_j/H_i)$, under the various hypotheses H_i , is available. Most systems do in fact allow a certain number of preliminary measurements in different real or simulated situations, from which histograms can be generated to get a numerical or analytical model of the distributions $p(s_j/H_i)$. The $I \times J$ values of probability density $p(s_j/H_i)$ associated respectively with the J local measurements s_j constitute the inputs for the processes discussed hereafter.

If we consider the most common case, where the measurements s_j can be assumed to be statistically independent, since the sensors are generally chosen for the complementary nature of the data they generate, the likelihood of each hypothesis H_i can be established immediately by the Bayesian approach, which typically calls for an evaluation of the *a posteriori* probability $P(H_i/s_1, \dots, s_J)$ of each hypothesis H_i using :

$$P(H_i/s_1, \dots, s_J) = \frac{\{\prod_j p(s_j/H_i)\} * P(H_i)}{\sum_k \{\prod_j p(s_j/H_k)\} * P(H_k)} \quad (2.10)$$

in which $P(H_i)$ designates its *a priori* probability.

However, this kind of approach quickly runs into difficulty when the real observation conditions differ from the available learning conditions, or when the measurement bank is not sufficient for a suitable learning process. The lack of control that can be seen at this level in most applications leads to distribution models that turn out to be more or less representative of the data actually encountered. In addition, it is often difficult to find a set of *a priori* probabilities $P(H_i)$ capable of reflecting the real situation with fidelity.

So we want to find a modeling based solely on the knowledge of $p(s_j/H_i)$ and capable of integrating any information concerning the reliability of the various

distributions, whether this comes from a more or less partial knowledge of the observations conditions or from a qualification of a data bank.

According to the generic approach introduced in section 2.1, any available qualitative information is assumed to be synthesized in the form of $I \times J$ coefficients $q_{ij} \in [0, 1]$, each being representative of a degree of confidence in the knowledge of each of the $I \times J$ distributions $p(s_j/H_i)$.

Dealing with this problem in the terms of evidence theory requires finding, for each source S_j , a model of its *I a priori* probabilities $p(s_j/H_i)$ and their *I* respective confidence factors q_{ij} in the form of a mass function $m_j(\cdot)$, characterized by a credibility function $Cr_j(\cdot)$, and by a plausibility function $Pl_j(\cdot)$. Since the sources S_j are distinct, a global evaluation $m(\cdot)$ can then be obtained by computation of the orthogonal sum of the $m_j(\cdot)$. The appropriate frame of discernment is of course the set of the *I a priori* listed hypotheses H_i .

To do this, we propose to conduct an exhaustive and exact search of all the models that might satisfy three fundamental axioms [2], [5]. These three axioms are chosen beforehand on the basis of their legitimacy in most of the applications concerned. They are :

Axiom 3 : Consistency with the Bayesian approach in the case where the learned distributions $p(s_j/H_i)$ are perfectly representative of the densities actually encountered ($q_{ij} = 1, \forall i, j$) and where the *a priori* probabilities $P(H_i)$ are known.

Axiom 4 : Separability of the evaluation of the hypotheses H_i ; that is, each probability must be considered as a distinct source of information generating a particular mass function $m_{ij}(\cdot)$, mainly capable of integrating the confidence factor q_{ij} specific to it. We thus require that each mass function $m_j(\cdot)$ be the orthogonal sum of the *I* mass functions $m_{ij}(\cdot)$ considered for $i \in [1, I]$. Also, considering the way the $p(s_j/H_i)$ probabilities are generated, the focal elements of the mass function $m_{ij}(\cdot)$ can be only H_i , $\neg H_i$, or E , where the frame of discernment E is the set of hypotheses H_i .

Axiom 5 : Consistency with the probabilistic association of the sources ; for independent sources S_j and densities $p(s_j/H_i)$ perfectly representative of reality, the modeling procedures retained must lead to the same result if we compute the orthogonal sum of the $m_j(\cdot)$ modeled from the $p(s_j/H_i)$ or if we model directly the joint probabilities $p(s_1, \dots, s_J/H_i)$ given by :

$$p(s_1, \dots, s_J/H_i) = \prod_j p(s_j/H_i) \quad (2.11)$$

The search for models satisfying these three axioms is presented in appendix B by progressively restricting the

set of possible models, taking the axioms into account in the order stated.

It appears that only two models satisfy the three desired axioms. Both meet the decomposition :

$$m_j(.) = \bigoplus_i m_{ij}(.) \quad (2.12)$$

Model 1 is particularized by :

$$m_{ij}(H_i) = 0 \quad (2.13)$$

$$m_{ij}(\neg H_i) = q_{ij} * \{1 - R_j * p(s_j/H_i)\} \quad (2.14)$$

$$m_{ij}(E) = 1 - q_{ij} + q_{ij} * R_j * p(s_j/H_i) \quad (2.15)$$

and Model 2 by :

$$m_{ij}(H_i) = q_{ij} * R_j * p(s_j/H_i) / \{1 + R_j * p(s_j/H_i)\} \quad (2.16)$$

$$m_{ij}(\neg H_i) = q_{ij} / \{1 + R_j * p(s_j/H_i)\} \quad (2.17)$$

$$m_{ij}(E) = 1 - q_{ij} \quad (2.18)$$

In both cases, the normalization factor R_j is simply constrained by :

$$R_j \in [0, (\max_{s_j, i} \{p(s_j/H_i)\})^{-1}] \quad (2.19)$$

Nevertheless, the specificity of the function used to generate model 2 allows R_j to be simply a positive number for this model in practice.

We may verify that this result is in fact a special case of the generic solution discussed in section 2.1. Both models provided by (2.3) to (2.5) and (2.6) to (2.8) in section 2.1 are strictly equivalent to the two models found here in (2.13) to (2.15) and in (2.16) to (2.18), if we adopt the following respective definitions for the C_{ij} :

$$\text{for model 1 : } C_{ij} = R_j * p(s_j/H_i) \quad (2.20)$$

$$\text{for model 2 : } C_{ij} = R_j * p(s_j/H_i) / [1 + R_j * p(s_j/H_i)] \quad (2.21)$$

in which R_j is still, of course, the normalization gain constrained by (2.19).

This outcome is in fact legitimate if we note that Axiom 1 is expressed directly by Axiom 4, and that the solutions required by Axioms 3 and 5 automatically verify Axiom 2. Axioms 3 and 5 simply make it possible to specify the inclusion of the particular information $p(s_j/H_i)$ in the expression of the criterion C_{ij} .

Lastly, when the data s_j are discrete values (local identity declarations, for example), the generalized Bayes theorem defined by P. SMETS in the framework of evidence theory [6] can be applied, for the case of statistical learning, to the cartesian product between the set of data and the set of hypotheses. It then strictly yields Model 1 developed here. Correlatively, we have

to note that the model 1, once again, minimizes the specificity measure criterion [4].

2.2.2 Uncertain Observation

We assume now that the measurements s_j are uncertain observations of discrimination features u_j , so that only their *a priori* probability distribution $p(s_j/u_j)$ is known. This may simply be the classical characterization of a measurement error. Furthermore, for each of the discrimination features u_j themselves, a learning of their *a priori* probability distribution $p(u_j/H_i)$ under the various hypotheses H_i is as formerly available.

If we note that :

$$p(s_j/H_i) = \int p(s_j/u_j) * p(u_j/H_i) du_j \quad (2.22)$$

then the results of section 2.2.1 can be directly extended so that the generic solution provided in section 2.1 holds with :

$$\text{for model 1 : } C_{ij} = R_j * \int p(s_j/u_j) * p(u_j/H_i) du_j \quad (2.23)$$

$$\text{for model 2 : } C_{ij} = R_j * [\int p(s_j/u_j) * p(u_j/H_i) du_j] / [1 + R_j * \int p(s_j/u_j) * p(u_j/H_i) du_j] \quad (2.24)$$

where the normalization factor R_j is now constrained by :

$$R_j \in [0, (\max_{s_j, i} \{ \int p(s_j/u_j) * p(u_j/H_i) du_j \})^{-1}] \quad (2.25)$$

2.2.3 Imprecise Observation

In this case the measurements s_j are imprecise observations of the discrimination features u_j , so that they only provide a fuzzy membership function $\mu_j(u_j)$. For each of the discrimination features u_j , a learning of their *a priori* probability distribution $p(u_j/H_i)$ under the various hypotheses H_i remains available, as formerly.

We can here express :

$$p(s_j/H_i) = \int \mu_j(u_j) * p(u_j/H_i) du_j \quad (2.26)$$

Once more the results of section 2.2.1 can be directly extended, and the generic solution provided in section 2.1 holds with :

$$\text{for model 1 : } C_{ij} = R_j * \int \mu_j(u_j) * p(u_j/H_i) du_j \quad (2.27)$$

$$\text{for model 2 : } C_{ij} = R_j * [\int \mu_j(u_j) * p(u_j/H_i) du_j] / [1 + R_j * \int \mu_j(u_j) * p(u_j/H_i) du_j] \quad (2.28)$$

The normalization factor R_j is now constrained by :

$$R_j \in [0, (\max_{s_j, i} \{ \int \mu_j(u_j) * p(u_j/H_i) du_j \})^{-1}] \quad (2.29)$$

2.3 Model With Approximate Prior Knowledge

Once again we have to consider the different kinds of relation between the sensor observations and the discrimination features, whereas the latter are now characterized by approximate prior knowledge.

2.3.1 Precise and Reliable Observation

In this case each of the measurements s_j is directly one of the discrimination features exploited, or a deterministic function of it. Nevertheless, the characterization of the different hypotheses H_i in the feature space is now provided by prior knowledge in the form of fuzzy membership functions $\mu_i(s_j)$. This means that an hypothesis H_i represents for instance an object about which we only know that it is large, slow, or heavy, as regards respectively the size, speed, or weight space.

To elaborate a suitable model, we have first to consider, for each hypothesis H_i , α -cuts $A_{\alpha k}$ of $\mu_i(s_j)$ at different decreasing levels α_k . As each α -cut defines a set that includes the previous one, it leads to the following consonant mass function on the measurement space :

$$m(A_{\alpha k}/H_i) = \alpha_k - \alpha_{k+1} \quad (2.30)$$

The Generalized Bayes Theorem defined by P. SMETS in the framework of the evidence theory [6] may then be implemented to obtain a mass function on E for each of the measurements s_j :

$$m(H_i/s_j) = 0 \quad (2.31)$$

$$m(\neg H_i/s_j) = 1 - \mu_i(s_j) \quad (2.32)$$

$$m(E/s_j) = \mu_i(s_j) \quad (2.33)$$

This mass function is then discounted at the rate $(1-q_{ij})$, if q_{ij} represents our degree of confidence in the prior knowledge $\mu_i(s_j)$, to provide the model we are looking for :

$$m_{ij}(H_i) = 0 \quad (2.34)$$

$$m_{ij}(\neg H_i) = q_{ij} * [1 - \mu_i(s_j)] \quad (2.35)$$

$$m_{ij}(E) = 1 - q_{ij} + q_{ij} * \mu_i(s_j) \quad (2.36)$$

Obviously this result can be expressed by the model 1 obtained for the generic problem in section 2.1, as soon as C_{ij} is defined by :

$$C_{ij} = \mu_i(s_j) \quad (2.37)$$

2.3.2 Uncertain Observation

We assume here that the measurements s_j are uncertain observations of discrimination features u_j , so that only their *a priori* probability distribution $p(s_j/u_j)$ is known. Furthermore, for each of the discrimination features u_j themselves, a prior knowledge provides the fuzzy membership function $\mu_i(u_j)$ that characterizes the different hypotheses H_i , as formerly.

On this basis, the possibility theory gives the possibility of each hypothesis H_i for each feature u_j :

$$\Pi(H_i/u_j) = \mu_i(u_j) \quad (2.38)$$

as well as the possibility density of each feature u_j for each measurement s_j :

$$\Pi(u_j/s_j) = R_j * p(s_j/u_j) \quad (2.39)$$

where R_j is a normalization factor such that :

$$R_j = [\max_{u_j} \{p(s_j/u_j)\}]^{-1} \quad (2.40)$$

As a possibility function is simply a consonant plausibility function, on the one hand (2.38) can in fact be deduced from (2.31), (2.32), and (2.33), reminding that in the latter s_j is directly the feature u_j , and on the other hand (2.39) can result from (2.12), (2.13), (2.14), and (2.15), when information is considered reliable ($q_{ij}=1$), if the set E of hypotheses becomes the infinite set of the possible values taken by u_j (note that such an evaluation in a continuous framework is only possible for the plausibility function, and thanks to the special nature of the focal elements of $m_{ij}(\cdot)$ as defined by (2.13), (2.14), and (2.15)). The normalization factor R_j has simply to be particularized by (2.40), in agreement with (2.19), in order to satisfy the definition of a possibility distribution.

Then, (2.38) and (2.39) allow to elaborate the possibility of each hypothesis H_i for each measurement s_j thanks to :

$$\Pi(H_i/s_j) = \sup_{u_j} \{ \Pi(H_i, u_j/s_j) \} \quad (2.41)$$

which can be developed as :

$$\Pi(H_i/s_j) = \sup_{u_j} \{ \Pi(H_i/u_j) \wedge \Pi(u_j/s_j) \} \quad (2.42)$$

where \wedge stands for the conjunction operator, and finally :

$$\Pi(H_i/s_j) = \sup_{u_j} \{ \min \{ \mu_i(u_j), R_j * p(s_j/u_j) \} \} \quad (2.43)$$

As a possibility function is but a particular plausibility function, the corresponding mass function which minimizes the specificity measure [4] can be expressed, after discounting according to the confidence factor q_{ij} :

$$m_{ij}(H_i) = 0 \quad (2.44)$$

$$m_{ij}(\neg H_i) = q_{ij} * [1 - \sup_{u_j} \{\min\{\mu_i(u_j), R_j * p(s_j/u_j)\}\}] \quad (2.45)$$

$$m_{ij}(E) = 1 - q_{ij} + q_{ij} * \sup_{u_j} \{\min\{\mu_i(u_j), R_j * p(s_j/u_j)\}\} \quad (2.46)$$

Once more, this result is obviously the model 1 provided for the generic problem in section 2.1, as soon as C_{ij} is now defined by :

$$C_{ij} = \sup_{u_j} \{\min\{\mu_i(u_j), R_j * p(s_j/u_j)\}\} \quad (2.47)$$

in which R_j is still given by (2.40).

2.3.3 Imprecise Observation

The measurements s_j are now imprecise observations of the discrimination features u_j , so that they only provide a fuzzy membership function $\mu_j(u_j)$. Nevertheless, for each of the discrimination features u_j themselves, the prior knowledge still provides the fuzzy membership function $\mu_i(u_j)$ that characterizes the different hypotheses H_i , as formerly.

The developments in these conditions are quite similar to those shown in section 2.3.2. The only difference concerns the expression of the possibility density of each feature u_j for each measurement s_j that becomes :

$$\Pi(u_j/s_j) = \mu_j(u_j) \quad (2.48)$$

Consequently, this leads once again to the model 1 provided for the generic problem in section 2.1, where C_{ij} has now to be defined by :

$$C_{ij} = \sup_{u_j} \{\min\{\mu_i(u_j), \mu_j(u_j)\}\} \quad (2.49)$$

2.4 Summary of the Models Obtained

A complete set of models has been developed in the previous sections, in the framework of the theory of evidence, according to the different kinds of data that have to be combined as regards measurements on the one hand, and prior knowledge on the other hand. All these models are particular cases of the two models provided for the generic problem in section 2.1, thanks to suitable expressions of C_{ij} . The definitions of C_{ij} for the different possible situations are summarized in table 1 for model 1 and in table 2 for model 2.

We may note that tables 1 and 2 furthermore provide expressions of C_{ij} for a precise and reliable prior knowledge, i.e. for a prior characterization of hypothesis H_i formalized by a deterministic value u_{ij} of the feature u_j . In fact these expressions are simply provided either by statistical learning when $p(u_j/H_i) = \delta_{u_j - u_{ij}}$, or by approximate prior knowledge when $\mu_i(u_{ij}) = 1$ and $\mu_i(u_j/u_j \neq u_{ij}) = 0$. Nevertheless we can verify the good coherence at this level between the two approaches in the common case of model 1.

Prior knowledge → Measurements ↓	u_{ij}	$p(u_j/H_i)$	$\mu_i(u_j)$
s_j	1 if $s_j = u_{ij}$ 0 if $s_j \neq u_{ij}$	$R_j * p(s_j/H_i)$ R_j constrained by (2.19)	$\mu_i(s_j)$
$p(s_j/u_j)$	$R_j * p(s_j/u_j)$ R_j constrained by (2.19)	$R_j * \int p(s_j/u_j) * p(u_j/H_i) du_j$ R_j constrained by (2.25)	$\sup_{u_j} \{\min\{\mu_i(u_j), R_j * p(s_j/u_j)\}\}$ u_j R_j defined by (2.40)
$\mu_j(u_j)$	$\mu_j(u_{ij})$	$R_j * \int \mu_j(u_j) * p(u_j/H_i) du_j$ R_j constrained by (2.29)	$\sup_{u_j} \{\min\{\mu_i(u_j), \mu_j(u_j)\}\}$ u_j

Table 1. Expression of C_{ij} in generic model 1 for the different kinds of prior knowledge and measurement

Prior knowledge → Measurements ↓	u_{ij}	$p(u_j/H_i)$	$\mu_j(u_j)$
s_j	0.5 if $s_j = u_{ij}$ 0 if $s_j \neq u_{ij}$	$R_j * p(s_j/H_i) / [1 + R_j * p(s_j/H_i)]$ $R_j \geq 0$	(*)
$p(s_j/u_j)$	$R_j * p(s_j/u_{ij}) / [1 + R_j * p(s_j/u_{ij})]$ $R_j \geq 0$	$R_j * [\int p(s_j/u_j) * p(u_j/H_i) du_j] /$ $[1 + R_j * \int p(s_j/u_j) * p(u_j/H_i) du_j]$ $R_j \geq 0$	(*)
$\mu_j(u_j)$	$\mu_j(u_{ij}) / [1 + \mu_j(u_{ij})]$	$R_j * [\int \mu_j(u_j) * p(u_j/H_i) du_j] /$ $[1 + R_j * \int \mu_j(u_j) * p(u_j/H_i) du_j]$ $R_j \geq 0$	(*)

(*) Not consistent

Table 2. Expression of C_{ij} in generic model 2 for the different kinds of prior knowledge and measurement

3 Target Classification

The target classification function consists in recognizing the type of a target, or even identifying it, on the basis of the different discriminating features s_j delivered by the sensors S_j that observe it. So the question is to designate the most likely hypothesis H_i^* in $E = \{H_1, \dots, H_I\}$ having regard to this information. Such a decision, which is immediate when a probability can be associated *a posteriori* with each hypothesis, becomes quite delicate when the evaluations are presented in terms of mass functions of the evidence theory. The whole difficulty revolves around the non-exclusivity of the evaluations, which raises the practical problem of interpretation and relative inclusion of the masses attached to those focal elements of cardinal 2 or greater, in the designation of a unique singleton. This problem, which is general to the evidence theory and unavoidable in the present context, has, as of today, been addressed only by more or less satisfactory intuitive solutions.

So below, we propose three different approaches to the problem of choosing the most likely hypothesis H_i^* , considering an arbitrary mass function $m(\cdot)$ on the frame of discernment $E = \{H_1, \dots, H_I\}$, when no other *a priori* basis for discriminating among the H_i is retained.

A synthesis of the resulting procedures provides a decision law suited to the classification problem. When applied to the modeled mass functions, this law supplies classification methods of noteworthy interest.

3.1 Minimum of Inconsistency

This approach consists in defining I certain mass functions $m_i(\cdot)$, each of them being respectively focused on each of the I hypotheses H_i of the frame of discernment E ($m_i(H_i) = 1$). The inconsistency K_i , provided by the orthogonal sum of the mass function $m_i(\cdot)$ and the available mass function $m(\cdot)$, reflects their disagreement, and so represents the conflict between the assessment $m(\cdot)$ and the fact that hypothesis H_i is

actually true. According to this, we have to choose the hypothesis H_i^* that ensures a minimal inconsistency K_i . As K_i can be written :

$$K_i = 1 - Pl(H_i) \quad (3.1)$$

in which $Pl(\cdot)$ is the plausibility function associated with $m(\cdot)$, we have to choose the hypothesis that provides a maximal plausibility.

The interest in this inconsistency criterion is confirmed by the idea of entropy that is connected with it [4].

3.2 « Bayesian » Approach

The idea here is to consider a given set of « equiprobable » Bayesian masses $m_0(\cdot)$ on the frame of discernment E ($m_0(H_i) = 1/I$, $\forall i \in [1, I]$). Endowing this mass function $m_0(\cdot)$ with a role similar to that of equiprobable *a priori* probabilities in the Bayesian inference, a mass function $m_c(\cdot)$ can be determined by orthogonal sum of the mass function $m_0(\cdot)$ and the available mass function $m(\cdot)$. $m_c(\cdot)$ is then a Bayesian mass function defined by :

$$m_c(H_i) = Pl(H_i) / \left\{ \sum_{k \in [1, I]} Pl(H_k) \right\}, \quad i \in [1, I] \quad (3.2)$$

$$m_c(A) = 0, \quad \forall A \neq H_i, \quad i \in [1, I] \quad (3.3)$$

in which $Pl(\cdot)$ is the plausibility function associated with $m(\cdot)$. By reference to the maximum *a posteriori* probability, the decision procedure obviously consists in retaining the hypothesis H_i^* that has the maximum mass, and thus the maximum plausibility here again.

Conceptually, the principle of this approach consists in substituting an equal confidence between the singletons of the frame of discernment in place of the total *a priori* uncertainty, so as to force the discrimination among these elements alone.

3.3 General Approach of the Decision Problem

Here we look for a solution with reference to a more general decisional context summarized, for example, in [7]. The purpose is to choose one of a number of possible actions to take, a_h , on the basis of the evaluation provided by the mass set $m(\cdot)$ on the frame of discernment E .

This choice can be made by maximizing a cost function $C(a_h)$ on the set of possible actions, knowing the weight $G(a_h/B_k)$ assigned to each potential action a_h when the event B_k , a subset of E , occurs :

$$C(a_h) = \sum_{B_k \subset E} \{G(a_h/B_k) * m(B_k)\} \quad (3.4)$$

The whole difficulty of using such a procedure in practice, and hence its credibility, resides in the evaluation of the weights $G(a_h/B_k)$, which is usually very subjective. While we may in general consider that the weights relative to the singletons H_i of E are given by the system or user, those relative to the subsets B_k of cardinal 2 or higher must, on the other hand, be determined intuitively, possibly in accordance with a preferred « attitude » [7].

Yet in our case, this subjective character can be greatly attenuated by the one-to-one correspondence we have to establish between the set of actions and the frame of discernment E , as each action a_h consists in declaring an hypothesis H_i to be true. So, if there are no specific requirements, the weights are legitimately given by :

$$G(a_i/B_k) = 1 \quad \text{if } H_i \in B_k \quad (3.5)$$

$$G(a_i/B_k) = 0 \quad \text{if } H_i \notin B_k \quad (3.6)$$

so as to conform with the associated idea of mass $m(B_k)$ as introduced by the evidence theory, i.e. as an evaluation of one of the elements of B_k , though we cannot specify of which element of B_k it is.

Under these conditions, (3.4) also leads to the designation of hypothesis H_i^* of maximum plausibility as the most likely.

Furthermore, it can be pointed out that this approach is coherent with the most consensual evaluation of the expected loss, among those proposed for instance in [8].

3.4 Synthesis

The three approaches presented all converge to the same decisional procedure, which consists in choosing the most likely hypothesis H_i^* according to :

$$H_i^* = \arg[\max_{i \in [1, J]} \{Pl(H_i)\}] \quad (3.7)$$

It should also be noted that this decision law is the one that satisfies the constraint emphasized at the end of the developments concerning axiom 3 in appendix B.

For all the models discussed in section 2, as the hypotheses H_i are singletons of the frame of discernment E , the plausibility $Pl(H_i)$ is proportional to the product of the $Pl_{kj}(H_i)$ associated with the $m_{kj}(\cdot)$, according to j and k . After normalization by the product of the $Pl_{kj}(\neg H_k)$ according to j and k , we come to designate H_i^* by the criterion :

$$H_i^* = \arg[\max_i \{ \prod_j ([m_{ij}(H_i) + m_{ij}(E)] / [m_{ij}(\neg H_i) + m_{ij}(E)]) \}] \quad (3.8)$$

This criterion can directly be applied to the two models provided for the generic problem, leading to the two respective solutions :

Solution 1 :

$$H_i^* = \arg[\max_i \{ \prod_j [1 - q_{ij} * (1 - C_{ij})] \}] \quad (3.9)$$

Solution 2 :

$$H_i^* = \arg[\max_i \{ \prod_j ([1 - q_{ij} * (1 - C_{ij})] / [1 - q_{ij} * C_{ij}]) \}] \quad (3.10)$$

It should be noted that solution 1 also meets a maximum credibility criterion.

The simplicity of the calculations and ease of use of these solutions is worth noting.

Furthermore, if we want to integrate a relative *a priori* confidence $\lambda_i \in [0, 1]$ respectively in the declaration of each of the different hypotheses H_i , or accordingly an expected risk $(1 - \lambda_i)$ attached to it, we may formalize this knowledge as a supplementary source of information. The plausibility function $Pl_0(\cdot)$ of such a subjective source has simply to be defined on E for the singletons H_i , the only information we need in the following, thanks to direct interpretation of the λ_i :

$$Pl_0(H_i) = \lambda_i \quad (3.11)$$

Therefore, the criterion of maximum plausibility (3.7) becomes, once the orthogonal sum between $Pl(\cdot)$ and $Pl_0(\cdot)$ is computed for the singletons H_i :

$$H_i^* = \arg[\max_{i \in [1, J]} \{\lambda_i * Pl(H_i)\}] \quad (3.12)$$

and consequently (3.9) and (3.10) become :

Solution 1 :

$$H_i^* = \arg[\max_i \{\lambda_i \prod_j [1 - q_{ij}^* (1 - C_{ij})]\}] \quad (3.13)$$

Solution 2 :

$$H_i^* = \arg[\max_i \{\lambda_i \prod_j [(1 - q_{ij}^* (1 - C_{ij})) / (1 - q_{ij}^* C_{ij})]\}] \quad (3.14)$$

Of course, in (3.9), (3.10), (3.13), and (3.14), C_{ij} stands for any of the suitable expressions developed in section 2, according to the nature of the available data.

3.5 Illustration

Figure 1 shows the mean probability of good recognition provided by the simulation of 2 sensors for a problem of discrimination between 2 hypotheses H_1 and H_2 . In this very simple example both sensors are similar, as regards either their *a priori* good discrimination capability, or the unreliability of their learning concerning hypothesis H_2 , in relation with an anticipated possible evolution of the context.

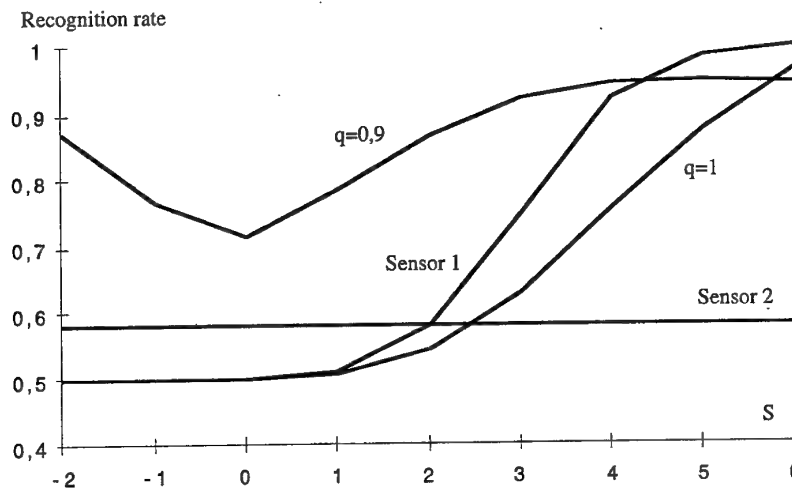


Figure 1. Classification with unreliable learning

More precisely, we are in the situation developed in section 2.2.1, and available learnings are given by normal distributions :

$$\begin{aligned} P(s_1/H_1) &= P(s_2/H_1) = N(0,1) \quad , \quad \text{with } q_{11} = q_{12} = 1 \\ P(s_1/H_2) &= P(s_2/H_2) = N(6,1) \quad , \quad \text{with } q_{21} = q_{22} = q \end{aligned}$$

while measurements actually simulated follow :

$$\begin{aligned} P(s_1/H_1) &= P(s_2/H_1) = N(0,1) \\ P(s_1/H_2) &= N(S,1) \quad , \quad P(s_2/H_2) = N(2,1) \end{aligned}$$

So in this test sensor 2 has effectively a wrong knowledge about H_2 , and the reliability of sensor 1 varies in function of the signal S due to H_2 . This is in accordance with the choice of factors q_{ij} that expresses a situation where a severe error concerning H_2 may occur simultaneously on both sensors. In this context our attention has to focus on the values of S much lower than 6, *i.e.* typically $S < 4$. Then the curves of figure 1 emphasize the robustness of our approach ($q=0.9$), as regards either the probabilistic approach, that is a

particular case of our method ($q=1$), or each sensor alone, which the probabilistic approach does not achieve.

Moreover, the aptitude of the q_{ij} factors to integrate linguistic or subjective information, considering the low sensitivity of the results to the choice of a given value for these coefficients, must be pointed out [5].

4 Management of Heterogeneous Frames of Discernment

Most often we have to use a sensor S_1 that provides a mass function $m_1(\cdot)$ on a frame of discernment E_1 , but in fact we need an assessment on a frame of discernment E such that $E_1 \subseteq E$, either for immediate decision making, or for combination with other mass functions available on complementary frames of discernment. For instance this may occur in case of incomplete learning, or when some hypotheses are not observable, as regards sensor S_1 . So we have to express $m_1(\cdot)$ on E .

The basic solution to that problem consists in deconditioning $m_1(\cdot)$ from E_1 to E (see appendix A). The advantage of this solution is that it does not require any particular condition, and therefore can be always implemented. Nevertheless, as we shall see, in most cases it is not the best approach.

A first alternative solution is possible when $m_1(\cdot)$ is obtained from modeling developed in section 2, *i.e.* it is the orthogonal sum, according to i , of the mass functions $m_{i1}(\cdot)$ defined by (2.3), (2.4), and (2.5), or by (2.6), (2.7) and (2.8). Then, the absence of any C_{i1} , characterized by $q_{i1}=0$, does amount to ignoring the corresponding mass function $m_{i1}(\cdot)$, as it becomes a trivial mass function ($m_{i1}(E)=1$), and is therefore a neutral element of the orthogonal sum. Moreover, the other elementary models $m_{i1}(\cdot)$ are not modified whatever the frame of discernment is, as their focal elements integrate in $\neg H_i$ any hypothesis that is different from H_i , so that $\neg H_i$ may naturally include all the missing hypotheses. Compared with the previous method of deconditioning at the level of $m_1(\cdot)$, this approach ensures less losses of specificity measure [4] as regards initial information.

If $m_1(\cdot)$ is directly used in a decision process, we have to point out that both approaches provide a maximal plausibility for missing hypotheses, as the latter have always a plausibility equal to 1. Decision rule (3.12) is then more suitable than decision rule (3.7), as λ_i can integrate the disparity in quantity of information that is available for each hypothesis.

Another kind of approach is presented hereafter, that is particular to situations requiring the orthogonal sum of several sources S_j defined on non disjoint respective frames of discernment E_j such that $\cup E_j = E$.

4.1 Plausibility Correction Method

This method (see for instance [9]) realizes a global treatment on the available sources, dealing with unlike frames of discernment and source combination together in a same processing. Nevertheless, to implement it, the sources to combine must be defined on frames directly or indirectly connected. It means that for a given source, there must exist at least another one such that their respective frames are not disjoint, thus having a common part.

Furthermore, this method is based on the use of some properties that the plausibility measure only verifies. Due to one of these properties, the method will be only able to deduce the plausibility of each hypothesis. In fact, there exists an infinity of mass functions that corresponds to this set of plausibilities. In particular, we can find the one that corresponds to the application of the minimum specificity criterion. Nevertheless, the relevant information remains based only on these few

plausibilities. Moreover, the maximum of plausibility is the decision criterion that is most often used in this theory, as justified in section 3. So, considering this criterion as the most suitable in accordance with the context introduced, the plausibility of each hypothesis is the only information that has to be expressed in the following.

Let S_1 and S_2 be two sources respectively defined on E_1 and E_2 , such that $E = E_1 \cup E_2$ and $E_c = E_1 \cap E_2$. The plausibility $Pl(H_i)$ of each hypothesis H_i after combination of both sources S_1 and S_2 can be rigorously expressed on the frame E as a function of the information actually available for each source in relation to their respective mass functions $m_1(\cdot)$ on E_1 and $m_2(\cdot)$ on E_2 . After suitable simplifications we obtain the following formulation of the relative plausibility of each hypothesis :

$$Pl(H_i) \equiv Pl_1(H_i), \quad \forall H_i \in (E_1 - E_c) \quad (4.1)$$

$$Pl(H_i) \equiv Pl_1(H_i) * Pl_2(H_i/E_c), \quad \forall H_i \in E_c \quad (4.2)$$

$$Pl(H_i) \equiv Pl_2(H_i) * Pl_1(E_c)/Pl_2(E_c), \quad \forall H_i \in (E_2 - E_c) \quad (4.3)$$

Because of the simplifications, the expressions provided are only proportional to plausibilities. Nevertheless, as the method is used in relation to a decision criterion of maximum of plausibility, the proportionality ratios of the plausibilities of the hypotheses are the only necessary information. Furthermore, symmetrical expressions can of course be obtained by permutation of the sources. Therefore, the decision is the same whatever the choice of development.

Intuitive explanation of the plausibility correction method may be the following. The method consists in choosing a reference source, and in refining and completing its knowledge by means of other sources. The reference source is S_1 in the present development, but this choice has no influence, as mentioned previously.

First, plausibilities of the hypotheses considered by S_1 and not common to the other source remain unchanged, which corresponds to the expression (4.1). Indeed, the other source gives no information about these hypotheses.

The second step consists in refining the knowledge the reference source has on the common hypotheses, thanks to a fusion with the other source on their common part E_c . Indeed, the latter represents the only subset on which it is legitimate to implement any fusion. This step corresponds to the expression (4.2). Obviously, such a step can exist only if there are at least two hypotheses contained in the common part E_c .

Lastly, knowledge of the reference source is completed by readjusting the plausibility of each hypothesis that is considered by S_2 and not by S_1 . So each of them is redefined relatively to the plausibilities of the hypotheses

of E_1 with respect to the common part E_c used as a pivot. This step simply consists in multiplying each of the plausibilities peculiar to S_2 by a unique factor $Pl_1(E_c)/Pl_2(E_c)$. Such a factor allows to preserve the proportionality ratio between the plausibilities of all of the hypotheses considered by the source S_2 . This operation corresponds to the expression (4.3).

Generalization of the plausibility correction method to more than two sources needs to determine an order of fusion, because this method is based on a non associative operation. The best approach consists in combining always sources that have the largest common part. With such an order, readjustments are based on the largest pivot, and so are more reliable. Furthermore, the maximum of hypotheses are involved in the fusion on common part. Nevertheless, for some configurations of sources, this order can still lead to several developments that are different from a decision point of view. Then it is necessary to determine the order that leads to the best performance, thanks to suitable criterion.

4.2 Introduction of Compatibility Relations

All the approaches that have been introduced can be extended in order to integrate further knowledge about some similarity that may exist between missing hypotheses in the initial frame of discernment and the hypotheses that are considered in it, as regards the features that are processed.

Such a similarity can be described by a compatibility relation ω_1 that associates to a considered hypothesis H_i of E_1 the set $\omega_1(H_i)$ of the hypotheses of $(E-E_1)$ to which H_i is similar, and that satisfies :

$$\omega_1(A) = \bigcup_{H_i \in A} \omega_1(H_i) \quad (4.4)$$

where $\omega_1(A)$ represents the set of missing hypotheses to which the considered hypotheses of E_1 contained in A are similar. Of course the approach imposes that every missing hypothesis is compatible with at least one of the hypotheses considered.

Deconditioning method can then be modified so that it consists now in appending (in the set union sense) to each focal element defined on E_1 only the missing hypotheses of $(E-E_1)$ with which this focal element is compatible. Doing so aims at considering that this element contains the hypotheses on which it is more plausible that the missing hypotheses discussed have transferred their evidence. Thus, the deconditioned mass function becomes :

$$m(A \cup \omega_1(A)) = m_1(A) , \quad \forall A \in N_1 \quad (4.5)$$

$$m(A) = 0 , \quad \forall A \in E, A \notin N_1 \quad (4.6)$$

where N_1 is the set of the focal elements of $m_1(\cdot)$.

Besides, the approach using modeling developed in section 2 can also be modified according to the same idea, providing on E , for each $H_i \in E_1$:

Model 1 :

$$m_{i1}(H_i \cup \omega_1(H_i)) = 0 \quad (4.7)$$

$$m_{i1}(\neg(H_i \cup \omega_1(H_i))) = q_{i1} * (1 - C_{i1}) \quad (4.8)$$

$$m_{i1}(E) = 1 - q_{i1} * (1 - C_{i1}) \quad (4.9)$$

Model 2 :

$$m_{i1}(H_i \cup \omega_1(H_i)) = q_{i1} * C_{i1} \quad (4.10)$$

$$m_{i1}(\neg(H_i \cup \omega_1(H_i))) = q_{i1} * (1 - C_{i1}) \quad (4.11)$$

$$m_{i1}(E) = 1 - q_{i1} \quad (4.12)$$

Then $m_1(\cdot)$ remains simply the orthogonal sum, according to i , of the mass functions $m_{i1}(\cdot)$ defined by (4.7), (4.8), and (4.9), or by (4.10), (4.11) and (4.12).

Concerning the plausibility correction method, some terms in the expression of the plausibility $Pl(H_i)$ obtain for each hypothesis H_i on the frame E , after combination of both sources S_1 and S_2 , increase in specificity as a source may now bring information on its missing hypotheses thanks to compatibility relations. After suitable simplifications the relative plausibility of each hypothesis becomes :

$$Pl(H_i) \equiv Pl_1(H_i) * Pl_2\left(\bigcup_{\omega_2(H_k) \cap H_i \neq \emptyset} H_k\right), \quad \forall H_i \in (E_1 - E_c) \quad (4.13)$$

$$Pl(H_i) \equiv Pl_1(H_i) * Pl_2(H_i/E_c), \quad \forall H_i \in E_c \quad (4.14)$$

$$Pl(H_i) \equiv Pl_2(H_i) * Pl_1\left(\bigcup_{\omega_1(H_k) \cap H_i \neq \emptyset} H_k\right) * Pl_1(E_c)/Pl_2(E_c), \quad \forall H_i \in (E_2 - E_c) \quad (4.15)$$

All the remarks emphasized in section 4.1 remain true.

4.3 Illustration

The deconditioning approach and the plausibility correction methods are applied on two sources S_1 and S_2 that are respectively defined on the frames $E_1 = \{H_1, H_2, H_3\}$ and $E_2 = \{H_2, H_3, H_4\}$. S_1 and S_2 are respectively SAR and infrared images of the ground, and the four classes to discriminate are : field (H_1), water (H_2), forest (H_3), building (H_4). The conditional probability densities of the features observed by each source with respect to each hypothesis of its frame are identified on actual images as normal distributions that are assumed to be perfectly representative. Consequently mass functions are built according to modeling provided in section 2.2.1, with $q_{ij} = 1$ for all i and j .

Processing of available images provides a confusion matrix for each source alone (tables 3 and 4), and for both fusion methods. An element of such a matrix is the declaration rate of an hypothesis H_k (column) when an hypothesis H_i is presented (row).

Declaration → Presentation ↓	H_1	H_2	H_3	H_4
H_1	.899	.690	.300	0
H_2	.492	.499	.800	0
H_3	.526	.750	.398	0
H_4	.837	.129	.340	0

Table 3. Confusion matrix of source 1

Declaration → Presentation ↓	H_1	H_2	H_3	H_4
H_1	0	.712	.320	.254
H_2	0	.925	.160	.590
H_3	0	.571	.224	.204
H_4	0	.108	.300	.889

Table 4. Confusion matrix of source 2

Figure 2 shows the recognition rates (diagonal elements of the confusion matrix) obtained for both fusion methods. The main difference is the advantage of the plausibility correction method as regards H_4 , thanks to a separate processing of H_1 , $H_2 \cup H_3$, and H_4 in this case.

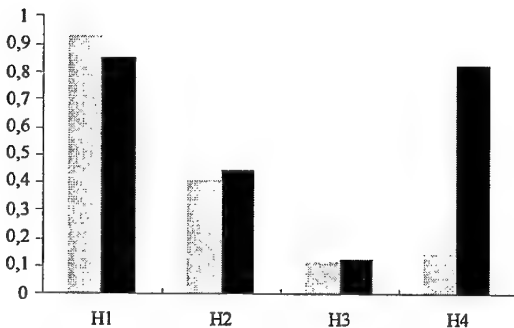


Figure 2. Recognition rate of deconditioning (gray) and plausibility correction method (black)

Figure 3 shows the same results when compatibility relations are implemented. In accordance with the individual confusion matrix of each source, we assume that H_4 is compatible with H_1 and H_2 as regards source 1, and that H_1 is compatible with H_2 and H_4 as regards source 2. The only difference with the previous

results concerns the better recognition of H_3 , thanks to compatibility relations that avoid any confusion between this hypothesis and the missing hypotheses of each source.

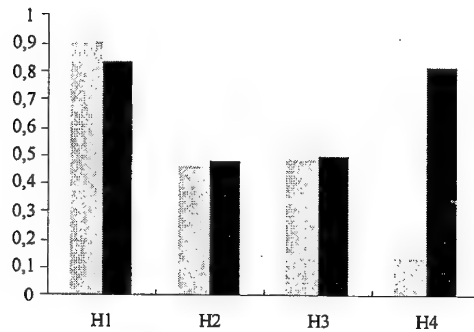


Figure 3. Recognition rate with compatibility relations (deconditioning in gray, plausibility correction in black)

5 Integration of Contextual Knowledge

Contextual information can obviously be integrated in all the previous processing methods thanks to the confidence factors q_{ij} that have been introduced to this end in the models elaborated in section 2. As pointed out in section 3.5, robust values of the q_{ij} can be arbitrarily determined beforehand for different possible situations to identify, from the moment sufficiently clear relations can be established between these situations and the effect of the confidence factors. When complex problems arise from that point of view, more efficient learning methods have to be implemented. Specific neural approaches have for instance been developed to this end [10].

Nevertheless, as the q_{ij} are continuous variables, their most efficient use consists in computing them on line on the basis of contextual parameters, provided that the latter are observable. Such a solution is presented in section 5.1, but considering the operational interest of adaptative processing based on contextual parameter measurements, the following introduces different solutions to this approach in a common framework [11, 12].

So we assume that a particular context $z = \{z_1, \dots, z_P\}$ is defined on a P -dimensional space Z by P contextual variables z_u that allow to evaluate the sensor reliability. Moreover, the vector $z^m = \{z_1^m, \dots, z_P^m\}$ represents the context measurements that are available.

The problem is to integrate the context measurements z^m in classification processes such as those developed in section 3, in order to improve their robustness to context variations. Two methods can be established to achieve such an integration. Both of them implement the same

combination rule between hypothesis assessments coming from sensor measurements on the one hand, and reliability information based on context measurements on the other hand, but at different levels. This rule is called the CC Rule (Contextual Combination Rule). Its presentation needs to define first :

- the inclusive validity domain D_V as the fuzzy subset of contexts ($D_V \subset Z$) in which a mass function $m_V(\cdot)$ is valid ;
- the index W as a subset of the set V of all the indexes v that characterize the different available mass functions $m_v(\cdot)$;
- the exclusive validity domain d_W as the fuzzy subset of contexts ($d_W \subset Z$) in which every mass function $m_v(\cdot)$ such that $v \in W$ is valid, but no other one ($v \notin W$) is valid :

$$d_W = \bigcap_{v \in W} D_v * \bigcap_{v \notin W} \neg D_v, \forall W \subseteq V - \{\emptyset\} \quad (5.1)$$

$$d_\emptyset = \bigcap_{v \in V} \neg D_v \quad (5.2)$$

In its general formulation the CC Rule consists in finding a global mass function $m(\cdot)$ on the frame of discernment $E = \{H_1, \dots, H_I\}$, considering on the one hand the mass functions $m_W(\cdot)$, each provided on E by the orthogonal sum of the mass functions $m_v(\cdot)$ such that $v \in W$, and on the other hand a mass function $m_c(\cdot)$ on $E_c = \{d_W\}$. The latter is assumed to be a Bayesian mass function which expresses the relative confidence we have in the different mass functions $m_W(\cdot)$.

First, a mass function $m'(\cdot)$ on $E_c \times E$ is obtained from the mass function $m_c(\cdot)$ relative to E_c and the mass functions $m_W(\cdot)$ relative to E , which has to be such that :

- the coarsening of $m'(\cdot)$ from $E_c \times E$ to E_c leads to the mass function $m_c(\cdot)$;
- the conditioning of $m'(\cdot)$ from $E_c \times E$ to $d_W \times E$ provides the mass function $m_W(\cdot)$.

These conditions are verified if :

$$m'(d_W, A) = m_c(d_W) * m_W(A), \forall A \subset E, \forall W \subseteq V \quad (5.3)$$

Then the final mass function $m(\cdot)$ on E is obtained by coarsening the mass function $m'(\cdot)$ from $E_c \times E$ to E :

$$m(A) = m_c(d_\emptyset) * m_\emptyset(A) + \sum_{W \subseteq V - \{\emptyset\}} m_c(d_W) * m_W(A), \quad \forall A \subset E \quad (5.4)$$

5.1 Local Contextual Combination Method (LCCM)

In this case a mass function $m_{cij}(\cdot)$ is associated with each elementary mass function $m_{ij}(\cdot)$ used by the models provided in section 2 (in which now $q_{ij}=1$), in order to characterize its reliability. $m_{cij}(\cdot)$ is established in the following on the frame of discernment $E_{cij} = \{D_{ij}, \neg D_{ij}\}$, where D_{ij} is the inclusive validity domain of $m_{ij}(\cdot)$.

Let the context $z = \{z_1, \dots, z_P\}$ be a random vector of probability density $p(z/z^m)$ where $z^m = \{z_1^m, \dots, z_P^m\}$ is the measurement vector associated to z . Besides, the validity domain D_{ij} is defined according to each contextual parameter z_u , in the framework of the fuzzy set theory, by an elementary membership function $\mu_{iju}(z_u)$. The membership function $\mu_{ij}(z)$ that characterizes the validity domain D_{ij} according to the context z is therefore expressed :

$$\mu_{ij}(z) = \min_u \{\mu_{iju}(z_u)\} \quad (5.5)$$

The probability $P(S_j/H_i, z^m)$ that the sensor S_j is reliable for the assessment of H_i if the context observation is z^m can then be obtained thanks to the definition of fuzzy event probability :

$$P(S_j/H_i, z^m) = \int \mu_{ij}(z) * p(z/z^m) dz \quad (5.6)$$

Of course, when the contextual variables are certain, the probability density $p(z/z^m)$ is replaced by the Dirac function $\delta(z - z^m)$, and (5.6) becomes :

$$P(S_j/H_i, z^m) = \mu_{ij}(z^m) \quad (5.7)$$

The probability (5.6) can in any case be finally formalized as a Bayesian mass function $m_{cij}(\cdot)$ such that :

$$m_{cij}(D_{ij}) = P(S_j/H_i, z^m) \quad (5.8)$$

$$m_{cij}(\neg D_{ij}) = 1 - P(S_j/H_i, z^m) \quad (5.9)$$

$$m_{cij}(E_{cij}) = 0 \quad (5.10)$$

Two mass functions $m_{Wij}(\cdot)$ ($W \in \{1, 2\}$) have now to be introduced : One of them uses the measurements as if they were completely reliable ($W=1$), while the other is representative of the total ignorance ($W=2$). These mass functions are therefore defined by :

$$m_{1ij}(\cdot) = m_{ij}(\cdot) \quad (5.11)$$

$$m_{2ij}(E) = 1 \quad (5.12)$$

Applying the CC Rule to the particular mass functions such that $m_c(\cdot)$ is $m_{cij}(\cdot)$ and $m_W(\cdot)$ are $m_{Wij}(\cdot)$ ($W \in \{1, 2\}$) provides a modified mass function $m'_{ij}(\cdot)$

that replaces the mass function $m_{ij}(\cdot)$ in the models developed in section 2 :

$$m'_{ij}(H_i) = P(S_j/H_i, z^m) * m_{ij}(H_i) \quad (5.13)$$

$$m'_{ij}(-H_i) = P(S_j/H_i, z^m) * m_{ij}(-H_i) \quad (5.14)$$

$$m'_{ij}(E) = 1 - P(S_j/H_i, z^m) + P(S_j/H_i, z^m) * m_{ij}(E) \quad (5.15)$$

The orthogonal sum of all the modified mass functions $m'_{ij}(\cdot)$ according to i and j provides finally the global mass function $m(\cdot)$ that is used in the decision rule, as explained in section 3.

In fact the resulting expression of $m'_{ij}(\cdot)$ is similar to the initial expression of $m_{ij}(\cdot)$ in which the confidence factor q_{ij} would be such that :

$$q_{ij} = P(S_j/H_i, z^m) \quad (5.16)$$

So, the LCCM is strictly the method developed in section 3, in which q_{ij} is simply expressed by (5.16), (5.6), and (5.5).

5.2 Global Contextual Processing Method (GCPM)

A unique mass function $m_c(\cdot)$ is now considered to characterize the relative validity of all the mass functions $m_{ij}(\cdot)$ and all their possible combinations. If W is a subset of $V = \{1, \dots, J\} \times \{1, \dots, J\}$, $m_c(\cdot)$ is more precisely defined on the frame of discernment $E_c = \{d_W\}$, provided that d_W stands for the fuzzy subset of contexts ($d_W \subset Z$) where all of the $m_{ij}(\cdot)$ such that $(i, j) \in W$ are valid, but only these ones.

The membership function $\mu_{ij}(z)$ that characterizes the validity domain D_{ij} of $m_{ij}(\cdot)$ according to the context z remains expressed as in section 5.1 by (5.5).

The probability of validity relative to the association W of mass functions $m_{ij}(\cdot)$ is the probability of the conjunction between their respective fuzzy subsets D_{ij} , when a contextual observation z^m is available :

$$P\left(\bigcap_{(i,j) \in W} D_{ij}/z^m\right) = \int [\min \{\mu_{ij}(z)\}] * p(z/z^m) dz \quad (5.17)$$

The exclusive probability of validity $P(d_W)$ relative to the association W of mass functions $m_{ij}(\cdot)$ can therefore be obtained on the basis of (5.17), thanks to developments similar to those provided in [13]. This probability is directly the Bayesian mass function $m_c(\cdot)$ that has to be expressed on E_c :

$$m_c(d_W) = \sum_{\substack{W' \subseteq V \\ W \subseteq W'}} (-1)^{|W'-W|} * P\left(\bigcap_{(i,j) \in W'} D_{ij}/z^m\right), \forall W \neq \emptyset \quad (5.18)$$

$$m_c(d_\emptyset) = P\left(\bigcap_{(i,j) \in V} \neg D_{ij}/z^m\right) \quad (5.19)$$

$|W'-W|$ represents the cardinal of the subset $W'-W$.

Correlatively, each mass function $m_W(\cdot)$ is the orthogonal sum of the mass functions $m_{ij}(\cdot)$ such that $(i, j) \in W$, if they exist :

$$m_W(\cdot) = \bigoplus_{(i,j) \in W} m_{ij}(\cdot), \quad \forall W \neq \emptyset \quad (5.20)$$

and $m_\emptyset(\cdot)$ corresponds to the absence of information :

$$m_\emptyset(E) = 1 \quad (5.21)$$

Thus, the CC Rule can be applied in its general form on (5.18), (5.19), (5.20), and (5.21), to provide the global mass function $m(\cdot)$ that is used in the decision rule, as explained in section 3 :

$$m(\cdot) = m_c(d_\emptyset) * m_\emptyset(\cdot) + \sum_{W \subseteq V - \{\emptyset\}} m_c(d_W) * m_W(\cdot) \quad (5.22)$$

5.3 Unified Formalism

A practical case of interest consists in adding a further hypothesis to E , corresponding to an unexpected objet for which no previous learning is available. This problem can be managed thanks to the models developed in section 2 that are used here, as explained in section 4. The modifications it involves in both methods, LCCM and GCPM, allow to express them in a same formalism.

Such a common formalism consists in implementing (5.20), (5.21), and (5.22) in the new extended frame of discernment. The difference between both methods appears in the expression of $m_c(\cdot)$ which remains given by (5.5), (5.17), (5.18), and (5.19) as regards GCPM, while LCCM has to use :

$$m_c(d_W) = \prod_{(i,j) \in W} q_{ij} * \prod_{(i,j) \notin W} (1 - q_{ij}) \quad (5.23)$$

$$m_c(d_\emptyset) = \prod_{(i,j) \in V} (1 - q_{ij}) \quad (5.24)$$

with q_{ij} defined by (5.16).

Furthermore, a numerical analysis shows that both approaches have complementary capacities when facing adverse situations, as regards the relative intrinsic efficiency and reliability of sensors, as well as the accuracy of the contextual observations.

5.4 Illustration

These methods have been implemented to deal with a problem of pixel fusion in multispectral image

processing, using the water vapor transmittance as contextual variable (see for instance [11]). This application involves two hypotheses (H_1 =Asphalt, H_2 =Vegetation), and two sensors ($S_1=2-2.3\mu\text{m}$, $S_2=0.4-0.6\mu\text{m}$). Previous learning of radiances, as regards a particular value of the water vapor transmittance, provides normal distributions of radiances specific to each sensor and each hypothesis as prior knowledge, so that the model developed in section 2.2.1 is used. Besides, every membership functions $\mu_{ij}(z)$ is determined beforehand thanks to the prior estimation of a physical relation between the water vapor transmittance and the mean radiance.

Figure 4 shows an example of mean recognition rates obtained by LCCM, probabilistic solution, and both sensors implemented alone, when the evolution of the water vapor transmittance is perfectly known. Results are presented as functions of the most sensitive mean radiance difference when the water vapor transmittance is decreasing from the learning condition to 0.

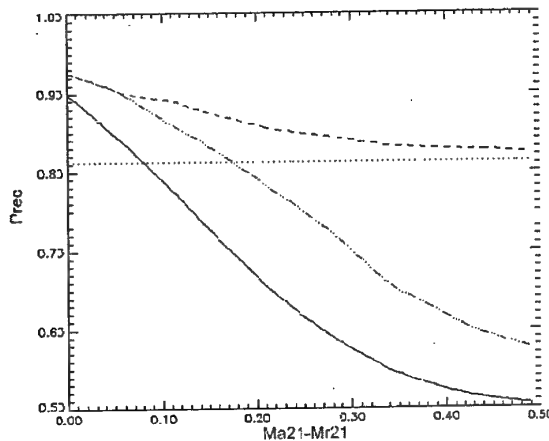


Figure 4. Recognition rate :
LCCM (---), probabilities (- . - . -),
sensor 1 (—), sensor 2 (.....)

These results emphasize the robustness brought by the adaptativity of LCCM.

6 Matching of Ambiguous Observations

Implementing sensors with orthogonal spatial resolutions, may improve the precision of target localization, thanks to the conjunction of local observations. It also provides a better separation between multiple targets in order to count them or to implement further analysis such as classification processes. For instance, such a situation occurs when we associate either delocalized passive sensors that provide local angle measurements, or an active sensor that provides distance and Doppler with a passive colocalized sensor that provides precise angular measurements.

Nevertheless, dealing with multiple targets in an area of interest leads to ambiguities in data association, due to ghosts (erroneous matching), hidden targets, non-detections, and false alarms, as illustrated in figure 5 in the case of two delocalized passive sensors.

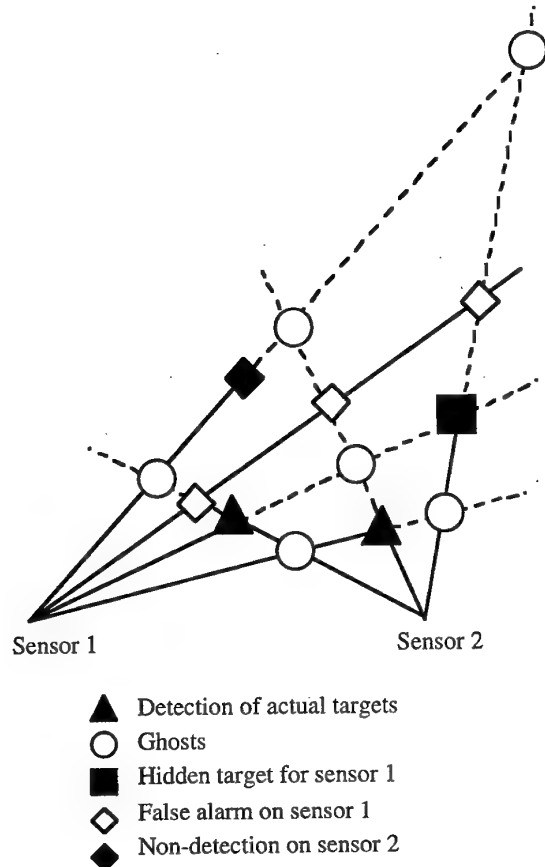


Figure 5. Ambiguous data association

To face this problem, a global approach of the detection, counting, localization, and classification functions is proposed in the following, that is a generalization of the developments provided in [14].

6.1 Formulation of the Problem

Developments are presented in the case of two sensors, in order to simplify expressions, but their generalization to any number of sensors is obvious, and the illustration in section 6.3 shows the implementation of 4 sensors.

Sensor 1 has N resolution cells x_1^n that are orthogonal to the M resolution cells x_2^m of sensor 2, as shown in figure 6. Furthermore x^{nm} stands for the intersection of cells x_1^n and x_2^m .

Each sensor provides a signal in each of its resolution cells : s_1^n is the signal provided by sensor 1 in cell x_1^n , and s_2^m is the signal provided by sensor 2 in cell x_2^m .

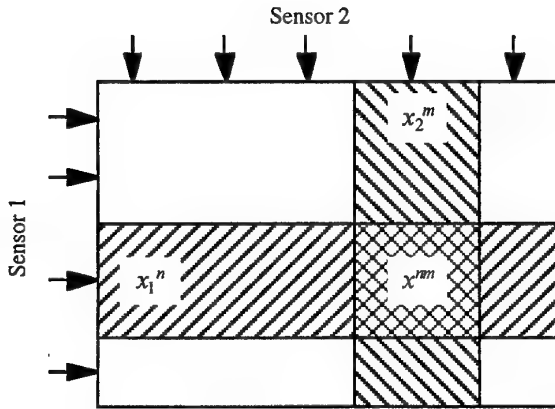


Figure 6. Definition of resolution cells

In each resolution cell x_1^n or x_2^m , there may be 0, 1, or several targets. According to the application encountered, a specific logic of target observability has to be considered. The following developments are valid for any of these logics, but for simpler explanation a particular one will be assumed. It is such that when there are several targets in a cell x_1^n (respectively x_2^m), the signal provided in this cell comes from the target that is the nearest from sensor 1 (respectively sensor 2). All the others are hidden by this first target, and consequently no information is available from sensor 1 (respectively 2) about their presence or identity. Nevertheless, it is assumed that there is at the most one target in an intersection x^{nm} .

On that basis, the process has to find the most likely distribution of targets, with their identity, over all the cells x^{nm} , as such a conclusion sums up the results of detection, counting, localization, and identification functions together. In other words, this consists in looking for the most likely singleton of :

$$E = \{(H_{11}, H_{21}, \dots, H_{N1}, H_{12}, \dots, H_{nm}, \dots, H_{NM})\} \quad (6.1)$$

where $H^{nm} = 0$ if there is no target in x^{nm} , and $H^{nm} = i$ if there is a target of identity i in x^{nm} .

To achieve such a purpose, three different types of information are considered :

- The likelihoods C_{i1}^n and C_{i2}^m of any hypothesis of identity i are respectively assessed in $[0,1]$ by sensors S_1 and S_2 on the basis of every signal s_1^n or s_2^m that each of them observes, and thanks to their suitable prior learning (probability distributions, fuzzy knowledge,...).

Furthermore, a confidence factor q_{i1}^n (respectively q_{i2}^m) with values in $[0,1]$ may be associated to each likelihood C_{i1}^n (respectively C_{i2}^m) in order to characterize its representativity in the context actually met. Moreover, the general assumptions made in section 2.1 about C_{ij} and q_{ij} are suitable as regards $[C_{i1}^n, q_{i1}^n]$ and $[C_{i2}^m, q_{i2}^m]$.

- The similarity R^{nm} of signals s_1^n and s_2^m is assessed on $[0,1]$, thanks to knowledge about physics and operational context that allows to characterize a similarity relation between signals coming from a same object (joint probability distribution, fuzzy relation,...). Once again, the general assumptions made in section 2.1 about C_{ij} are suitable for R^{nm} , as the latter is a trivial case of the generic problem : 1 source (S_1 =comparator), 2 hypotheses (H_1 =similarity, H_2 =-similarity), all $q_{ij}=1$, and incomplete information (no C_{21} available).

- A logic of matching is defined to integrate the logic of target observability introduced above. Such a logic leads simply to a suitable definition of the frames of discernment and focal elements that are used to model any available information. Considering the specific logic of observability introduced above as example, the likelihood C_{i1}^n has to express that s_1^n allows discrimination on :

$$E_1^n = \{H_{01}^n, H_{11}^n, \dots, H_{i1}^n, \dots, H_{I1}^n\} \quad (6.2)$$

where H_{01}^n means « no target in x_1^n », and H_{i1}^n means « at least one target in x_1^n , and the identity of the nearest one is i », for every i from 1 to I . Similarly, the likelihood C_{i2}^m has to express that s_2^m allows discrimination on :

$$E_2^m = \{H_{02}^m, H_{12}^m, \dots, H_{i2}^m, \dots, H_{I2}^m\} \quad (6.3)$$

where H_{02}^m means « no target in x_2^m », and H_{i2}^m means « at least one target in x_2^m , and the identity of the nearest one is i », for every i from 1 to I .

Furthermore, the similarity of two signals s_1^n and s_2^m means in this specific case that there is a target in x^{nm} , and no other target in front of it, either for sensor 1 or for sensor 2 (see assumption about hidden targets). So, a similarity analysis between s_1^n and s_2^m allows discrimination on $\{A^{nm}, P^{nm}\}$ which is a partition of E defined by :

$$Pnm = \{H^{11}, \dots, H^{ij}, \dots, H^{NM}\}, \text{ with :}$$

$$\begin{aligned} H^{ij} &\neq 0 \text{ if } i=n \text{ and } j=m \\ H^{ij} &= 0 \text{ if } i=n \text{ and } j \neq m \\ H^{ij} &= 0 \text{ if } i \neq n \text{ and } j=m \\ \text{Either } H^{ij} &= 0 \text{ or } H^{ij} \neq 0 \text{ if } i \neq n \text{ and } j \neq m \\ \text{Either } H^{ij} &= 0 \text{ or } H^{ij} \neq 0 \text{ if } i=n \text{ and } j>m \\ \text{Either } H^{ij} &= 0 \text{ or } H^{ij} \neq 0 \text{ if } i>n \text{ and } j=m \end{aligned} \quad (6.4)$$

$$A^{nm} = E \cdot P^{nm} \quad (6.5)$$

Figure 7 shows a representation of Pnm .

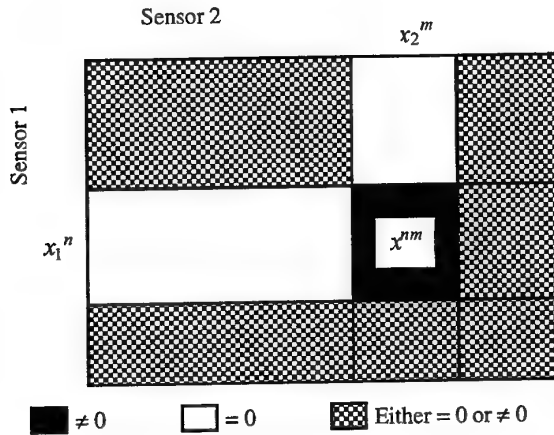


Figure 7. Description of Pnm

6.2 General solution

A first mass function $m^\circ(\cdot)$ has to be elaborated on the basis of the « classification information », i.e. all the likelihoods C_{i1}^n and C_{i2}^m and their confidence factor. The frame of discernment of $m^\circ(\cdot)$ must be E , as regards the purpose of the process.

Nevertheless, according to assumptions of the problem, the $I+1$ likelihoods C_{i1}^n attached respectively to each of the $I+1$ hypotheses of identity H_{i1}^n that correspond to a same resolution cell x_1^n of sensor 1, and their associated confidence factor, have to generate a distinct mass function $m_1^n(\cdot)$ on their specific frame of discernment E_1^n . Then, the mass functions $m_1^n(\cdot)$ and their counterparts $m_2^m(\cdot)$ for sensor 2 must be refined from their own frame of discernment E_1^n or E_2^m to the common frame E in order to allow their orthogonal sum and provide the expected mass function $m^\circ(\cdot)$.

The formulation of each $m_1^n(\cdot)$ or $m_2^m(\cdot)$ as a function of the corresponding C_{i1}^n or C_{i2}^m and the associated confidence factors has to fit in with the developments provided in section 2.1, so that it is the orthogonal sum

of mass functions $m_{i1}^n(\cdot)$ (respectively $m_{i2}^m(\cdot)$), according to i , which are expressed by the less specific model (2.3), (2.4), and (2.5), implemented on C_{i1}^n and q_{i1}^n (respectively C_{i2}^m and q_{i2}^m) for $J=1$. Furthermore, all expressions of C_{ij} provided in section 2 according to the nature of learning and observation (precise value, probability distribution, or fuzzy membership function) are suitable for C_{i1}^n and C_{i2}^m .

A second mass function $m^{\circ\circ}(\cdot)$ has also to be elaborated on the basis of the « similarity information », i.e. the similarity R^{nm} considered for all the hypotheses of signal association (s_1^n, s_2^m) , having regard to matching logic. The frame of discernment of $m^{\circ\circ}(\cdot)$ must be E , as regards the purpose of the process.

According to the definition of the similarity relations, each hypothesis of signal association (s_1^n, s_2^m) has to be considered separately, so that a mass function $m^{nm}(\cdot)$ is first established on $\{A^{nm}, Pnm\}$ thanks to the less specific generic model expressed by (2.3), (2.4), and (2.5) in section 2.1, as regards the equivalence between R^{nm} and the generic problem that has been emphasized in section 6.1. $m^{\circ\circ}(\cdot)$ is then simply the orthogonal sum of the mass functions $m^{nm}(\cdot)$, according to n and m .

The orthogonal sum of $m^\circ(\cdot)$ and $m^{\circ\circ}(\cdot)$ leads finally to a global mass function $m(\cdot)$ on E that allows to select the expected most likely singleton of E , thanks to decision making rules presented in section 3.

More precisely, the process consists in the seven following steps :

- Step 1 : for the $N \times M$ resolution cells of both sensors, elaboration of the $(I+1)$ mass functions $m_{i1}^n(\cdot)$ or $m_{i2}^m(\cdot)$ that model the classification information as introduced above. The corresponding plausibility functions, that are necessary in the following, are respectively :

$$Pl_{i1}^n(H_{i1}^n) = 1 - q_{i1}^n + q_{i1}^n \cdot C_{i1}^n \quad (6.6)$$

$$Pl_{i1}^n(\neg H_{i1}^n) = 1 \quad (6.7)$$

and :

$$Pl_{i2}^m(H_{i2}^m) = 1 - q_{i2}^m + q_{i2}^m \cdot C_{i2}^m \quad (6.8)$$

$$Pl_{i2}^m(\neg H_{i2}^m) = 1 \quad (6.9)$$

reminding that all expressions of C_{ij} provided in section 2 according to the nature of learning and observation (precise value, probability distribution, or fuzzy membership function) are suitable for the elaboration of C_{i1}^n and C_{i2}^m .

- Step 2 : Refining each of these $(I+1) \times (N+M)$ mass functions from E_1^n (respectively E_2^m) in E .

- *Step 3* : Orthogonal sum of the $(I+1)*(N+M)$ mass functions in E . As a result of steps 2 and 3, the plausibility function corresponding to the resulting mass function $m^\circ(.)$ can be expressed simply for the singletons of E (only these terms are necessary in the following) :

$$Pl^\circ(H^{11}, \dots, H^{nm}, \dots, H^{NM}) = (1-K^\circ)^{-1} * \prod_n \prod_i Pl_{i1}^n(X_{i1}^n) * \prod_m \prod_i Pl_{i2}^m(X_{i2}^m) \quad (6.10)$$

with :

$$X_{01}^n = H_{01}^n \text{ if } \forall m, H^{nm}=0$$

$$X_{01}^n = \neg H_{01}^n \text{ if } \exists m, H^{nm} \neq 0$$

$$X_{i1}^n = H_{i1}^n, \text{ for } 1 \leq i \leq I, \text{ if } \exists m : H^{nm}=i$$

$$\forall n' < n, H^{n'm}=0$$

$$\forall m' < m, H^{nm'}=0$$

$$X_{i1}^n = \neg H_{i1}^n, \text{ for } 1 \leq i \leq I, \text{ otherwise}$$

$$X_{02}^m = H_{02}^m \text{ if } \forall n, H^{nm}=0$$

$$X_{02}^m = \neg H_{02}^m \text{ if } \exists n, H^{nm} \neq 0$$

$$X_{i2}^m = H_{i2}^m, \text{ for } 1 \leq i \leq I, \text{ if } \exists n : H^{nm}=i$$

$$\forall n' < n, H^{n'm}=0$$

$$\forall m' < m, H^{nm'}=0$$

$$X_{i2}^m = \neg H_{i2}^m, \text{ for } 1 \leq i \leq I, \text{ otherwise}$$

and in which K° is the combination inconsistency, the expression of which is not useful for the following.

- *Step 4* : Elaboration of the $N*M$ mass functions $m^{nm}(.)$ that model the similarity information as introduced above. The corresponding plausibility functions, that are necessary in the following, are respectively :

$$Pl^{nm}(P^{nm}) = R^{nm} \quad (6.11)$$

$$Pl^{nm}(A^{nm}) = 1 \quad (6.12)$$

- *Step 5* : Orthogonal sum of these $N*M$ mass functions in E . The plausibility function corresponding to the resulting mass function $m^\circ(.)$ can be expressed simply for the singletons of E (only these terms are necessary in the following) :

$$Pl^\circ(H^{11}, \dots, H^{nm}, \dots, H^{NM}) = (1-K^{\circ\circ})^{-1} * \prod_n \prod_m Pl^{nm}(X^{nm}) \quad (6.13)$$

with :

$$X^{nm} = P^{nm} \text{ if } H^{nm} \neq 0$$

$$H^{kl}=0 \text{ for } k=n \text{ and } l < m$$

$$H^{kl}=0 \text{ for } k < n \text{ and } l=m$$

$$X^{nm} = A^{nm} \text{ otherwise}$$

and in which $K^{\circ\circ}$ is the combination inconsistency, the expression of which is not useful for the following.

- *Step 6* : Orthogonal sum of $m^\circ(.)$ and $m^{\circ\circ}(.)$ in E to provide the resulting mass function $m(.)$. The corresponding plausibility function can be expressed for the singletons of E :

$$Pl(H^{11}, \dots, H^{NM}) = (1-K)^{-1} * Pl^\circ(H^{11}, \dots, H^{NM}) * Pl^{\circ\circ}(H^{11}, \dots, H^{NM}) \quad (6.14)$$

in which K is the combination inconsistency, the expression of which is not useful for the following.

- *Step 7* : Selection of the singleton(s) that provide(s) a maximum of the plausibility function $Pl(.)$, according to the decision principles discussed in section 3.

6.3 Illustration

The implementation of a very simple example is described, in order to discuss how this method works. Four sensors are observing a same area, according to the situation illustrated in figure 8. Each sensor has only two resolution cells, and these cells are the same ones, on the one hand between sensors 1 and 3 :

$$x_1^1 = x_3^1 = x^{11} \cup x^{12}$$

$$x_1^2 = x_3^2 = x^{21} \cup x^{22}$$

and on the other hand between sensors 2 and 4 :

$$x_2^1 = x_4^1 = x^{11} \cup x^{21}$$

$$x_2^2 = x_4^2 = x^{12} \cup x^{22}$$

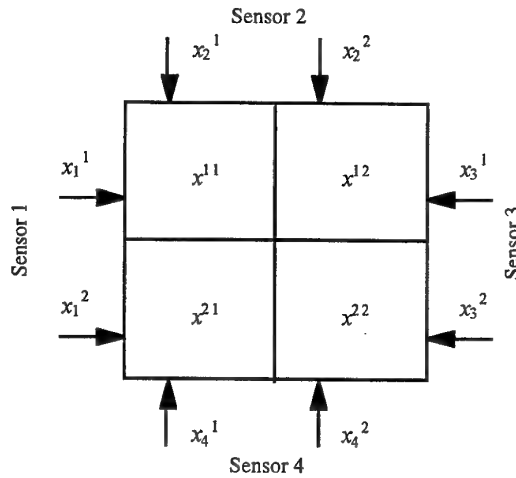


Figure 8. Implementation of sensors

The use of four sensors ensures that a target will always be seen by at least two sensors with orthogonal resolution cells in any circumstances, in spite of the phenomenon of hidden target. So, in principle, the localization of all the targets is possible. The only

problems that may remain concern the classification, and mainly the ambiguities in matching, on which we want to focus. Furthermore, it is assumed that there is only one possible identity of target, so that the problem of classification is reduced to a problem of detection (absence or presence of a target).

The signal in each cell of each sensor is an energy measurement generated thanks to random variables which have a normal distribution $N(0,1)$ if there is no target in the cell, and $N(3,1)$ if there is at least one target in it. Furthermore, the variables that simulate two different signals are independent if the signals are not coming from a same target, and the standard deviation of their difference is constrained to 0.1 if the signals are coming from a same target.

In this numerical application the probability distributions of signals are learned without any problem of representativity, so that, for the two cells h of any of the four sensors k , $q_{ik}^h=1$ and the likelihoods C_{ik}^h are given by the model (2.20) on the basis of :

$$p(s_k^h/H_{0k}^h) = N(0,1)$$

$$p(s_k^h/H_{1k}^h) = N(3,1)$$

According to their definition (2.19), the corresponding normalization factors are :

$$R_k = (2*\pi)^{1/2}$$

To complete prior knowledge, the similarity relation between signals coming from a same target expresses that these signals are « almost equal », in accordance with the correlation that is simulated between actual signals (see above). Therefore, the set of associations $(s_k^h, s_{k'}^{h'})$ of signals coming from a same object is assumed to be a fuzzy subset of the set of all the possible associations, that is characterized by the membership function $\mu(s_k^h, s_{k'}^{h'})$ dedicated to cell h of sensor k and cell h' of sensor k' . $\mu(s_k^h, s_{k'}^{h'})$ is drawn in figure 9 as a function of $|s_k^h - s_{k'}^{h'}|$.

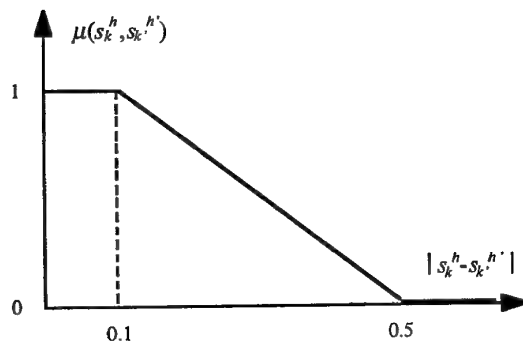


Figure 9. Similarity relation between s_k^h and $s_{k'}^{h'}$

Simulations have been run for the 16 possible actual distributions of targets over the 4 cells x^{nm} , including all numbers and positions of targets. A statistical analysis of results is presented in figure 10. It provides the mean recognition rate of the right target distribution (singleton of E), for all the possible actual distributions, as a function of a maximum number of ambiguities, i.e. the rate of the right distribution recognition when we allow to declare at the most the number of ambiguities given by abscissa. Ambiguities are multiple declarations of target distributions, including the right one, when these distributions lead to the same maximum value of criterion. Note that it is different from the confusion notion that characterizes wrong declarations.

Results are shown for three different criteria :

- *Classification criterion* : it consists in maximizing $PI^c(\cdot)$, which uses only classification information.

- *Similarity criterion* : it consists in maximizing $PI^s(\cdot)$, which uses only similarity information.

- *Global criterion* : it consists in maximizing $PI(\cdot)$, which uses all the available information (classification + similarity). It corresponds to the method that has been elaborated in previous sections.

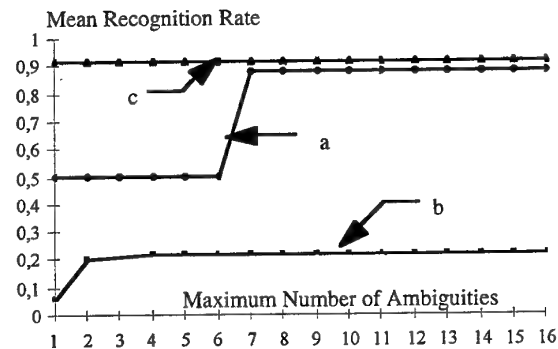


Figure 10. Recognition rate of target distributions

a - Classification criterion

b - Similarity criterion

c - Global criterion

These results emphasize several behaviors :

- as expected, the method achieves a good recognition of target distributions (i.e. at the same time a good detection, counting and localization), without any ambiguity.

- the two types of information processed (classification and similarity) show a very good complementarity, as the result of their association provides a much better performance than any of them alone.

- classification information has a quiet good recognition capacity, but it suffers from a number of ambiguities that reduces drastically its performance.

- similarity information has a very poor recognition capacity, and furthermore it suffers from ambiguities. Nevertheless it allows to reject all the ambiguities of classification information, and at the same time it improves the recognition rate of the latter.

7 Target Tracking

The problem dealt with here is that of tracking a moving target of any possible nature, in a dense environment, using observations delivered by a set of disparate and possibly delocalized sensors. One of the main purposes is to overcome the problem of spurious sources present in the vicinity of the target. These sources may be due to intelligent countermeasures, artifacts, or vehicles that are untracked, for operational or technical reasons. A situation of major practical interest appears when the tracking is initialized on objects that are very close together or even at the same point, such as when a fighter plane enters an airspace hidden behind or close to an airliner. Simultaneous tracking of multiple targets may also be suitably handled with the proposed approach, as is shown by the extension proposed in section 7.2.

Unlike classical methods, the concept proposed performs a filtering directly on the discriminatory features available in the different resolution cells of each sensor, rather than on plots provided by a detection procedure [15]. So, it is elaborated according to a global approach that integrates in a same processing both tracking and classification functions.

Although it constitutes no particular limitation on the concept emphasized, the discussion here presumes that the target tracked is the only one of its particular identity in the space being processed, and that a given resolution cell contains at most one target of any given identity.

The technique used for the filtering aspects is inspired directly from the Probabilistic Data Association Filter (PDAF) family of methods developed by Y. BAR SHALOM from the ordinary KALMAN filter, to handle multiple detections [16]. These methods differ essentially from the KALMAN filter by the estimate updating phase, in which they proceed in two steps :

- First the statistical gating selects the detected plots located in a given vicinity of the predicted position. The vicinity is determined so as to contain the target with an *a priori* probability greater than a given threshold.
- Then the estimate and its covariance are updated on the basis of an innovation determined by linear combination of the innovations individually due to each plot retained as potential successor of the processed track. The weighting coefficients are the *a priori* probabilities for each of these plots to actually be due to

the target, considering the detection and false alarm probabilities of the detector used, the predicted position and its covariance, and the statistical gating threshold.

In a first approximation, the method proposed here can be interpreted as a PDAF whose detection would operate at minimum threshold, with Detection Probability = False Alarm Probability = 1. At the level of the statistical gating, then, this is equivalent to retaining and processing one plot per resolution cell located within the vicinity defined around the predicted position.

The « *a priori* » probability that weights the innovation due to each of these plots in updating the estimate is, on the other hand, modified to reflect the likelihood of the identity present in the corresponding cell, information generated thanks to the recognition of identity features extracted from the signal isolated by the spatial resolution of the sensors.

The expression for the filtering, prediction, and statistical gating modules specifically for this modified version of the PDAF is given in the appendix C. The following discussion concerns the special development of the cell weightings then necessary for the innovation, that requires two indispensable ideas to be defined :

- The sensors are said to be « aligned » if they break the validation gate down into the same resolution cells. For convenience here, the sensors are assumed to be classed in groups of sensors that are aligned among themselves, while two sensors of two different groups are necessarily unaligned. Each sensor will thus be denoted S_j^l , where l designates to which of the L groups of aligned sensors the sensor in question belongs, and j is its sequence number within the group of J sensors.

- If, for a group l of aligned sensors, x^{ln} designates the n th of N resolution cells of non-empty intersection with the validation gate, then the sensors in question « resolve » the gate if the gate entirely includes each x^{ln} .

7.1 Procedure Description

The extraction of features in each resolution cell x^{ln} by each sensor S_j^l is assumed to provide information of the type considered by the generic model developed in section 2, and therefore by any of the models proposed in this section. So we have $I \times J$ mass functions $m_{ij}(\cdot)$ per resolution cell x^{ln} , with each of them being defined either by (2.3), (2.4), and (2.5), or by (2.6), (2.7), and (2.8). So they are, from now on, denoted $m_{ij}^{ln}(\cdot)$, by reference to the resolution cell x^{ln} to which each of them relates, and their respective frames of discernment are the corresponding partitions of E denoted $E_i^{ln} = \{H_i^{ln}, -H_i^{ln}\}$, where the hypothesis H_i^{ln} means that identity H_i is present in cell x^{ln} . It will be noted that the use of the models established above is advantageous in light of their suitability to the problems

generally encountered, but that this is not indispensable : the discussion here starts with any given mass functions $m_{ij}^{ln}(\cdot)$, which can be obtained by any other means.

The procedure therefore consists in combining the various sources $m_{ij}^{ln}(\cdot)$, each being specific to a sensor S_j^l , a resolution cell x^{ln} , and a particular evaluated identity H_i . The combination is performed in such a way as to provide the likelihood of each possible distribution of identity hypotheses (including target absence) on the M resolution cells x^m of the validation gate. The x^m cells are the intersections of the x^{ln} cells of the various groups l of sensors, so that the combination processes applied offer the best spatial resolution at the end of the process.

Thanks to a special property of evidence theory, the combination of the resulting likelihoods with the *a priori* localization probabilities (α^0, α^m) of the tracked target delivered by the filter prediction, directly generates the *a posteriori* localization probabilities (β^0, β^m) of the

target in question. The probabilities α^m and β^m are relative to the presence of the target in the cell x^m (hypothesis H^m), while the probabilities α^0 and β^0 concern its absence in the gate (hypothesis H^0). The probabilities β^m and β^0 are used to weight the innovation due to each of the cells x^m in the estimate update, as was introduced above.

Considering the nature of the problem, the required combinations must be performed by orthogonal sum of all of the sources, to obtain their conjunction. This must be done in the finest common frame of discernment, which is the set E^F of the possible identity distributions on the various cells x^m of the validation gate. As the orthogonal sum is commutative, the association order of the various sources is theoretically arbitrary. To simplify the calculations, however, the approach chosen consists in associating the sources by order of decreasing similarity of their frame of discernment, whereas applying appropriate refinements at each step. Figure 11 shows the resulting logic of operations.

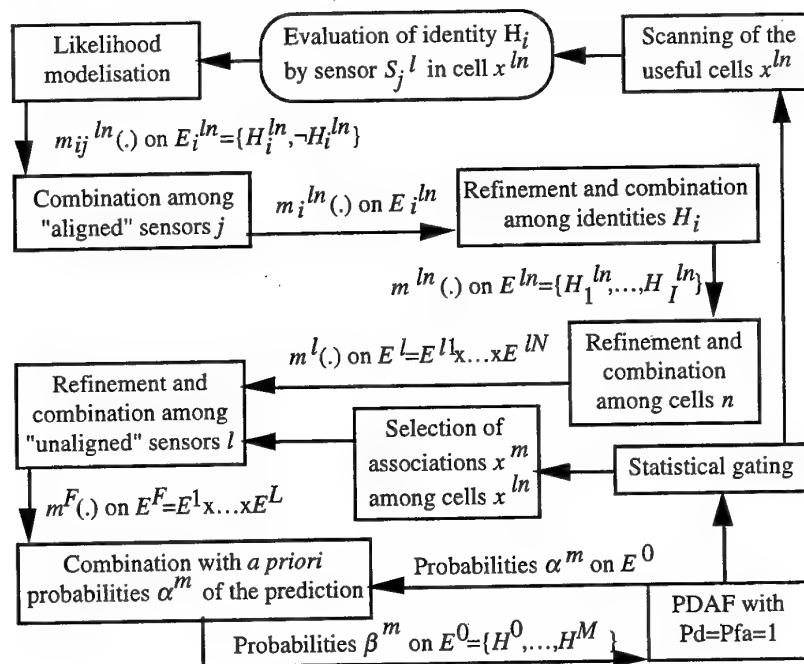


Figure 11. Combination processing

Developments following this scheme are provided in appendix D. They lead to the expression of the probabilities β^m and β^0 that the filter detailed in appendix C requires :

$$\beta^0 = \alpha^0 / \{ \alpha^0 + \sum_{m=1}^M \alpha^m * Q^m \} \quad (7.1)$$

$$\beta^m = \alpha^m * Q^m / \{ \alpha^0 + \sum_{m'=1}^M \alpha^{m'} * Q^{m'} \} \quad (7.2)$$

$$\text{in which : } Q^m = \prod_{l=1}^L Q^{ln} \quad (7.3)$$

$$x^m \subset x^{ln}$$

with, for the sensor groups l that resolve the gate :

$$Q^{ln} = \prod_{j=1}^J A_{ij}^{ln} / \{ 1 - \prod_{i=1}^{I-1} (1 - \prod_{j=1}^J B_{ij}^{ln}) + \sum_{i=1}^{I-1} (\prod_{j=1}^J A_{ij}^{ln} - \prod_{j=1}^J B_{ij}^{ln}) \} \quad (7.4)$$

and, for the sensor groups l that do not resolve the gate :

$$Q^{ln} = \prod_{j=1}^J A_{ij}^{ln} / \{ 1 - \prod_{i=1}^I (1 - \prod_{j=1}^J B_{ij}^{ln}) + \sum_{i=1}^I (\prod_{j=1}^J A_{ij}^{ln} - \prod_{j=1}^J B_{ij}^{ln}) \} \quad (7.5)$$

In both cases, H_I designates the identity of the tracked target, and the coefficients A_{ij}^{ln} and B_{ij}^{ln} represent, respectively, the expressions :

$$A_{ij}^{ln} = \{ m_{ij}^{ln}(H_I^{ln}) + m_{ij}^{ln}(E_i^{ln}) \} / \{ m_{ij}^{ln}(-H_I^{ln}) + m_{ij}^{ln}(E_i^{ln}) \} \quad (7.6)$$

$$B_{ij}^{ln} = m_{ij}^{ln}(E_i^{ln}) / \{ m_{ij}^{ln}(-H_I^{ln}) + m_{ij}^{ln}(E_i^{ln}) \} \quad (7.7)$$

The resulting filter will hereafter be designated the Multiple Signal Filter (MSF).

Let us note that, if we consider only two identity hypotheses (*i.e.* absence or presence of a target in each cell in question), if the available information is of the probabilistic type used by the models developed in section 2.2.1, and if the distributions are perfectly representative of reality (all $q_{ij}^{ln}=1$), then there exists a Bayesian solution to the problem. It is easy to verify that this solution actually is the special case of the filter proposed for the discussed conditions. Such a solution was, for example, used in the PDAFAI to include the amplitude of the observed signal [17].

On the other hand, as soon as the number of identity hypotheses exceeds 2 (absence/presence of the tracked target, in the cell considered), no formal probabilistic approach is possible any more, since the prediction can no longer provide the *a priori* probabilities of the different identities needed for the Bayesian inference of the update. One of the advantages of the approach proposed is therefore to obtain an exact solution for these situations, which are especially of concern here (see introduction to section 7). The method described also makes it possible to manage the uncertainty on the models and to include data that is not necessarily probabilistic.

This concept also naturally provides track validation criteria, typically based on the likelihood of the actual presence of a target of the desired identity in the validation gate, in consideration of the various features observed.

It should also be noted that such a filter is by nature suited to a given identity, the purpose of the proposed concept being to reject as effectively as possible those signals that might be due to neighboring targets of different identity. In track initialization phase, a battery of different filters suitable for different identities should therefore be used. The filter whose identity is most likely can be chosen progressively using track validation criteria. This organization also makes it possible to adopt the most appropriate dynamic model for the identity processed, for each filter.

Furthermore, in the particular case of unaligned sensors, similarity relations such as introduced in section 6 can be integrated in the present processing, in order to improve matching of ambiguous observations. The weights Q^m are then simply multiplied by a complementary factor $Pl^{oc}(H^m)/Pl^{oc}(H^0)$, where $Pl^{oc}(\cdot)$ is the plausibility associated with the mass function $m^{oc}(\cdot)$ defined at the step 5 of the process described in section 6.2. This holds because H^m and H^0 , specific subsets of E , are entirely included in every focal element that supports it, whatever the initial mass function considered here may be.

7.2 Joint Tracking of Multiple Targets

We now propose to extend the single-target concept above to the joint tracking of multiple targets whose validation gates overlap. The purpose is therefore to develop a new Joint Multiple Signal Filter (JMSF) from the Joint Data Association Filter (JDAF) of Y. BAR SHALOM [16], using the approach that allowed us to establish the MSF from the PDAF. Let P be the number of tracks concerned. All the notations used up to now are conserved, with an added subscript p to indicate the track to which the notation refers.

The formulation sought can be obtained by refining the mass functions $m_p^F(\cdot)$, defined at the level of each track (see figure 11), from E_p^F to the cartesian product $EX=E_p^F \times \dots \times E_p^F$, and performing their orthogonal sum in this new frame of discernment. The result should then be conditioned and coarsened on the cartesian product of the $E_0^p = \{H_0^p, \dots, HM_p^p\}$, minus the different target position combinations in which more than one target is located in a same resolution cell. The mass function obtained can then be combined with the weighting coefficients $\alpha^m p_p$ that would be used in a JPDAF operating at Detection Probability = 1, similar to those found in (C1) and (C2) of appendix C for the PDAF. These coefficients are actually identical to *a priori* probabilities in the frame of discernment considered. This leads us to the *a posteriori* weighting coefficients

β^{mp}_p needed for updating the filters associated, respectively, with each track p :

$$\beta^{mp}_p = D^{-1} * \sum_{\substack{m_p \in [0, M_p] \\ p' \in [1, P] - \{p\} \\ x_{m1} \neq \dots \neq x_{mP}}} \{ \alpha^{mp}_p * P l_p(H^{mp}_p) * \prod_{p' \in [1, P] - \{p\}} [\alpha^{mp'}_{p'} * P l_{p'}(H^{mp'}_{p'})] \} \quad (7.8)$$

where, by convention, $m_p=0$ corresponds to a position of target p outside the gate, and where D is the normalization factor that guarantees :

$$\sum_{m_p=0}^{M_p} \beta^{mp}_p = 1 \quad (7.9)$$

In practice, (7.8) is therefore expressed :

$$\beta^{mp}_p = D^{-1} * \alpha^{mp}_p * Q^{mp}_p * \sum_{\substack{m_p \in [0, M_p] \\ p' \in [1, P] - \{p\} \\ x_{m1} \neq \dots \neq x_{mP}}} \prod_{p'} \alpha^{mp'}_{p'} * Q^{mp'}_{p'} \quad (7.10)$$

with, for each track p , $Q^0_p=1$ and Q^{mp}_p is given by (7.3) to (7.7).

7.3 Illustration

Two very simple simulations are used to illustrate some of the potential advantages of MSF [15], with reference to the most suitable classical method. The latter consists in an usual PDAF, associated with a classification before tracking that aims at declaring at first the identity present in each resolution cell, thanks to a Maximum Likelihood criterion. It will be noted PDAF^C.

The only attribute considered in each resolution cell is the observed signal level, with this level being characterized by its *a priori* probability density under the various possible identity hypotheses : $N(0,1)$ for no target, and $N(S,1)$ when one is present, with S being able to take different values depending on the target considered. So, the modeling developed in section 2.2.1 is used

The trajectories simulated are straight and level, at constant speed, approaching the sensors colocalized in (0,0) head-on. The dynamic model used in the filter is the same as the one that generates the trajectories. The only error introduced at the level of the filter concerns the track position and velocity initializations. The real trajectories are in dotted lines and the estimated ones in solid lines.

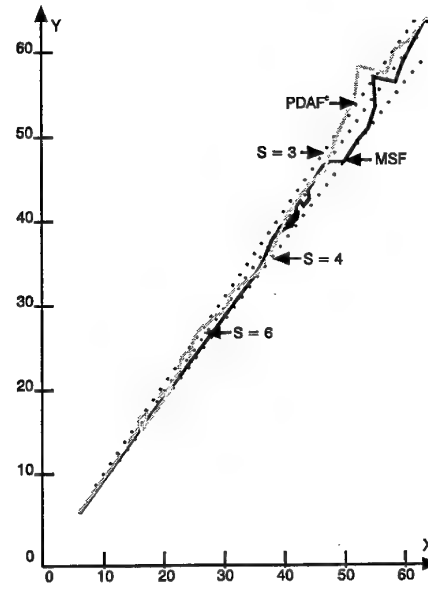


Figure 12. Tracking with a 2D radar in (0,0)

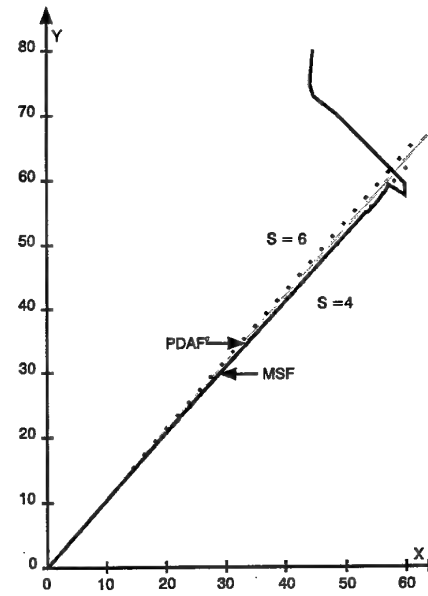


Figure 13. Tracking with 1 pulse radar and 1 optonic imager in (0,0)

In figure 12 three targets ($S=3$, $S=4$, $S=6$) are observed by a 2D radar (azimuth and range), and the one we are trying to track is target $S=4$, which is, therefore, hemmed in between two targets : one weaker in signal power and the other stronger. Under these particularly difficult conditions, the MSF converges much faster, and on the right target, while the PDAF^C can only lock onto the more powerful one, hampered by the unavoidable limitations of his detection phase. This emphasizes the inability of the PDAF^C to meet the compromise

between convergence and rejection of spurious sources as well as FMS does it.

In figure 13 radar range measurements are associated with azimuth measurements from an optronic imager. Each of the two targets present is assumed to induce the same signal level from the two sensors, which is respectively $S=4$ and $S=6$, and the target to be tracked is $S=4$. In this context plot processing induces a ghost phenomenon due to misclassifications of target $S=6$, i.e. association ambiguities between the range measurements and azimuth measurements, of the kind that set off persistent false alarms in the PDAFC, which therefore generates a trajectory lying along the plot of barycenters between the two real trajectories. The MSF on the other hand, quickly locks onto the right target and tracks it correctly, thanks to its ability to match the measurements better according to the identities of the targets that originate them, and thereby to reject the incorrect identities better.

8 Conclusion

The theory of evidence proves to be an interesting federative framework for multisensor processing, as it allows to integrate data and information of disparate nature, thanks to appropriate modeling that has been elaborated. On this basis, suitable processings have been developed to achieve the main functions required in situation assessment, such as classification, matching of ambiguous observations, and tracking.

Their major advantage is a better robustness in adverse conditions, thanks to their ability to manage uncertainty, unreliability, and incomplete knowledge. They also allow to get the best out of the available information thanks to global functional approaches and centralized upstream data fusion, i.e. where data are the most informative, what they are able to achieve in any case.

Furthermore, they provide useful capabilities for multisensor systems implementation, such as management of heterogeneous frames of discernment or integration of contextual knowledge, in order to get the best out of complementary available sensors and ensure suitable adaptativity.

References

1. G. Shafer : A Mathematical Theory of Evidence. Princeton University Press, Princeton, New Jersey, 1976.
2. A. Appriou : Uncertain Data Aggregation in Classification and Tracking Processes. In « Aggregation and Fusion of Imperfect Information », B. Bouchon-Meunier, ed., Physica Verlag, 1998.
3. A. Appriou : Multisensor Data Fusion in Situation Assessment Processes. In « Qualitative and Quantitative Practical Reasoning », D. Gabbay, R. Kruse, A. Nonnengart, H. J. Ohlbach, eds, Publ. Springer, Lecture Notes in Artificial Intelligence 1244, 1997.
4. R.R. Yager : Entropy and Specificity in a Mathematical Theory of Evidence. International Journal General Systems, Vol. 9, 1983, pp 249-260.
5. A. Appriou : Probabilités et incertitude en fusion de données multi-senseurs. Revue Scientifique et Technique de la Défense, n°11, 1991-1, pp 27-40.
6. P. Smets : Belief Functions. In « Nonstandard Logics for Automated Reasoning », P. Smets, A. Mamdani, D. Dubois, H. Prade, eds, Academic Press, London, 1988.
7. R.R. Yager : A General Approach to Decision Making with Evidential Knowledge. Uncertainty in Artificial Intelligence, L.N. Kanal & J.F. Lemmer eds, Elsevier Science Publishers, B.V. North-Holland, 1986.
8. T. Dencœur : Analysis of Evidence-Theoretic Decision Rules for Pattern Classification. Pattern Recognition, Vol. 30, N° 7, pp 1095-1107, 1997.
9. F. Janez, A. Appriou : Theory of Evidence and non Exhaustive Frames of Discernment - Plausibilities Correction Method. International Journal of Approximate Reasoning 18 (1998), Elsevier.
10. M.C. Perron-Gitton : Apport d'une approche neuro-floue dans un contexte de fusion de données basé sur la théorie de l'évidence. IPMU' 94, Paris, 4-8 juillet 1994.
11. S. Fabre, A. Appriou, X. Briottet, P. Marthon : Pixel Fusion - Contribution of Contextual Physical Data for the « a Priori » Database Construction. First International Symposium on Physics in Signal and Image Processing, PSIP'99, Paris, January 18-21, 1999.
12. S. Fabre, A. Appriou, X. Briottet : Comparison of two Integration Methods of Contextual Information in Pixel Fusion. Second International Conference on Information Fusion, FUSION'99, Sunnyvale, CA, USA, July 6-8, 1999.
13. V. Nimier : Introduction d'informations contextuelles dans des algorithmes de fusion multicapteur. Traitement de Signal, Vol. 14, n° 5, Spécial 1997.
14. A. Appriou : Spatially Ambiguous Multisensor Data Processing. First International Conference on Multisource-multisensor Information Fusion, FUSION'98, Las Vegas, Nevada, USA, July 6-9, 1998.
15. A. Appriou : Multiple Signal Tracking Processes. Aerospace Science and Technology, n° 2, February 1997.
16. Y. Bar Shalom, T.E. Fortmann : Tracking and Data Association. Academic Press, New York, 1988.
17. D. Lerro, Y. Bar Shalom : Interacting Multiple Model Tracking with Target Amplitude Feature. IEEE Transactions on Aerospace and Electronic Systems, Vol. 29, n° 2, april 1993.

Appendix A: Theory of Evidence, Useful Basic Considerations

The theory of evidence starts from the definition of a frame of discernment E including I exclusive hypotheses H_i ($i \in [1, I]$). 2^E is then the set of the $2^I - 1$ subsets of E .

A few basic functions defined from 2^E to the interval $[0, 1]$ allow to characterize the likelihood of any subset of E :

- the mass function $m(\cdot)$, that represents the likelihood of the singletons belonging to a subset assessed, without possible discernment between these singletons. It is such that:

$$\sum_{A \subseteq E} m(A) = 1 \quad (A1)$$

$$m(\emptyset) = 0 \quad (A2)$$

- the credibility function $Cr(\cdot)$, which may be interpreted as a kind of minimal likelihood of a subset, and which is bound to the mass function thanks to:

$$Cr(B) = \sum_{A \subseteq B} m(A) \quad (A3)$$

- the plausibility function $Pl(\cdot)$, which may be interpreted as a kind of maximal likelihood of a subset, and which is bound respectively to the mass function and to the credibility function thanks to:

$$Pl(B) = \sum_{A \cap B \neq \emptyset} m(A) \quad (A4)$$

$$Pl(B) = 1 - Cr(\neg B) \quad (A5)$$

- the commonality function $Q(\cdot)$, which is defined as:

$$Q(B) = \sum_{A \subseteq E, B \subseteq A} m(A) \quad (A6)$$

The focal elements of a mass function $m(\cdot)$ are the elements A of 2^E such that $m(A)$ is not null. When the focal elements are reduced to the singletons H_i of the frame of discernment E , then the mass function $m(\cdot)$ is a Bayesian mass function, which is identical to the credibility and the plausibility functions, all of them being identical to the classical notion of probability. When the focal elements are all included in one another, then the credibility and plausibility functions are respectively reduced to the notions of necessity and possibility used in the possibility theory.

Combination of distinct sources

The most legitimate combination rule, according to an appropriate axiomatic description, is the conjunction provided by the orthogonal sum, which allows to aggregate J distinct mass functions $m_j(\cdot)$ ($j \in [1, J]$) defined on a same frame of discernment E :

$$m(\cdot) = m_1(\cdot) \oplus m_2(\cdot) \oplus \dots \oplus m_J(\cdot) \quad (A7)$$

and which is such that:

$$m(A) = (1-K)^{-1} \sum_{A_1 \cap \dots \cap A_J = A \neq \emptyset} \prod_{j=1}^J m_j(A_j) \quad (A8)$$

where A_j stands for any focal element of $m_j(\cdot)$, and K is the inconsistency of the fusion, *i.e.* the degree of conflict between the different mass functions:

$$K = \sum_{A_1 \cap \dots \cap A_J = \emptyset} \prod_{j=1}^J m_j(A_j) \quad (A9)$$

This rule can be expressed very simply on the basis of the respective commonality functions:

$$Q(A) = (1-K)^{-1} \prod_{j=1}^J Q_j(A) \quad (A10)$$

Thanks to the definitions provided above, (A10) can be applied to the plausibilities of singletons H_i .

Nevertheless, the implementation of the orthogonal sum is meaningful only if the inconsistency is not too important, *i.e.* if there is no major conflict between the different sources. Such a conflict may arise in different circumstances, for instance when sources are not reliable, when the frame of discernment is not exhaustive, or when sources are not assessing the same object because of spatially ambiguous observations. According to the problem encountered, different attitudes are possible. First, instead of using the inconsistency for normalization, it may be assigned to the empty set, to the whole frame of discernment, or to a further singleton H_{I+1} added to E with deconditioning (see below) on the new frame. Another way consists in implementing a disjunction:

$$m(A) = \sum_{A_1 \cup \dots \cup A_J = A} \prod_{j=1}^J m_j(A_j) \quad (A11)$$

or an adaptative law between conjunction and disjunction:

$$m(A) = \sum_{A_1 \cap \dots \cap A_J = A} \prod_{j=1}^J m_j(A_j) + \sum_{A_1 \cup \dots \cup A_J = A} \prod_{j=1}^J m_j(A_j) \quad (A12)$$

$$A_1 \cap \dots \cap A_J = \emptyset$$

Different convex combinations are also possible, either between conjunction and disjunction, or directly between source assessments.

All these laws are generally commutative (except the convex combinations), but several are not associative (nonetheless the orthogonal sum is associative). Furthermore, when applied to probabilities or possibilities, which are particular cases of plausibility functions, they generally do not provide outputs of the same nature (except the orthogonal sum applied to probabilities).

Reliability management

The lack of reliability of a source is managed by discounting its mass function $m(\cdot)$ at level d . This operation provides an updated mass function $m^d(\cdot)$ such that :

$$m^d(A) = (1-d)*m(A), \quad \forall A \subseteq E, A \neq E \quad (A13)$$

$$m^d(E) = d + (1-d)*m(E) \quad (A14)$$

Management of frames of discernment

Four basic operations allow to manage both granularity (refining / coarsening) and exhaustivity (conditioning / deconditioning) of frames of discernment :

(i) A *refining* R associates to each hypothesis H_i^1 of a frame of discernment $E^1 = \{H_1^1, \dots, H_{I_1}^1\}$ a subset $R(H_i^1)$ of another frame $E^2 = \{H_1^2, \dots, H_{I_2}^2\}$, such that $\{R(H_1^1), \dots, R(H_{I_1}^1)\}$ is a partition of E^2 . So a mass function $m^1(\cdot)$ defined on E^1 provides a mass function $m^2(\cdot)$ on E^2 thanks to an operation of minimal extension :

$$m^2(R(A)) = m^1(A), \quad \forall A \subseteq E^1 \quad (A15)$$

(ii) A *coarsening* C is the inverse operation of a refining R ($C=R^{-1}$), so that a mass function $m^1(\cdot)$ on E^1 is obtained from a mass function $m^2(\cdot)$ on E^2 thanks to :

$$m^1(A) = \sum_{B \subseteq E^2} m^2(B) \quad (A16)$$

$$A = \{H_i^1 / R(H_i^1) \cap B \neq \emptyset\}$$

(iii) *Conditioning* allows to specify an assessment by the introduction of a further information. More precisely, the certainty of a proposition A is expressed by a mass function $m_A(\cdot)$ such that :

$$m_A(A) = 1 \quad (A17)$$

and the conditioning of an initial mass function $m(\cdot)$ according to this proposition consists in combining it with $m_A(\cdot)$:

$$m(\cdot|A) = m(\cdot) \oplus m_A(\cdot) \quad (A18)$$

(iiii) *Deconditioning* consists in obtaining, from a mass function $m_1(\cdot)$ defined on E_1 , a mass function $m(\cdot)$ on a frame of discernment E broader than E_1 ($E_1 \subseteq E$). $m(\cdot)$ must be such that its conditioning on E_1 is exactly $m_1(\cdot)$. Among all the mass functions that satisfy this condition, $m(\cdot)$ is selected according to the principle of minimum assignment :

$$m(A \cup (E - E_1)) = m_1(A), \quad \forall A \subseteq E_1 \quad (A19)$$

Appendix B : Statistical Learning Modeling

The search for all the models satisfying the three axioms proposed in section 2.2.1 is led by progressively restricting the set of possible models.

Axiom 3 : Consistency With the Bayesian Approach

Development

Let $m_0(\cdot)$ be the mass function representative of information source S_0 consisting of the *a priori* probabilities $P(H_i)$. $m_0(\cdot)$ is then a Bayesian mass function defined by :

$$m_0(H_i) = P(H_i), \quad \forall i \in [1, I] \quad (B1)$$

$$m_0(A) = 0, \quad \forall A \neq H_i, i \in [1, I] \quad (B2)$$

The desired consistency requires that the orthogonal sum of the mass functions $m_j(\cdot)$ and $m_0(\cdot)$ produces a Bayesian mass function $m_b(\cdot)$ in conformity with the Bayesian inference (2.10) whenever the distributions $p(s_j/H_i)$ are perfectly representative of the densities actually encountered, and thus whenever $q_{ij}=1$ for any i and j . This axiom should, in particular, remain true for any subset of combined sources S_j delimited by $j \in J' \subseteq [1, J]$. Concretely :

$$m_b(\cdot) = \left\{ \bigoplus_{j \in J'} m_j(\cdot) \right\} \oplus m_0(\cdot) \quad (B3)$$

should under these conditions therefore verify :

$$m_b(H_i) = \{ [\prod_j p(s_j/H_i)] * P(H_i) \} / \sum_k \{ [\prod_j p(s_j/H_k)] * P(H_k) \}, \quad \forall H_i \in E \quad (B4)$$

Moreover, equations (B1), (B2), and (B3) lead to :

$$m_b(H_i) = \{ [\prod_j Pl_j(H_i)] * P(H_i) \} / \sum_k \{ [\prod_j Pl_j(H_k)] * P(H_k) \}, \quad \forall H_i \in E \quad (B5)$$

By satisfying (B4) and (B5) jointly for any $J' \subset [1, I]$ we lastly define each $m_j(\cdot)$ by its plausibility function using the I equations :

$$Pl_j(H_i) = K_j * p(s_j/H_i), \quad i \in [1, I] \quad (B6)$$

in which K_j is a unique parameter for the I equations, defined simply by :

$$K_j \in [\{ \sum_i p(s_j/H_i) \}^{-1}, \{ \max_i [p(s_j/H_i)] \}^{-1}] \quad (B7)$$

These bounds on K_j are required only by the intrinsic nature of the idea of plausibility, which has to remain less than unity, while the sum of the values it takes for events constituting a partition of E (the H_i themselves, here) must be greater than unity.

Comments

The conclusion thus drawn from Axiom 3 calls for a few comments. Firstly, in the general case where $I > 2$, for each value of K_j other than the minimum required by (B7), there exists an infinite number of possible mass functions, defined by a system of $I+1$ equations (I equations (B6) and the sum of the masses equal to 1) with $2I-1$ unknowns.

For the minimum value of K_j , the result obtained always amounts to a unique, and moreover Bayesian, mass function :

$$m_j(H_i) = p(s_j/H_i) / \sum_k p(s_j/H_k), \quad \forall i \in [1, I] \quad (B8)$$

$$m_j(A) = 0, \quad \forall A \neq H_i, \quad i \in [1, I] \quad (B9)$$

Of the various solutions obtained for the maximum value of K_j , there exists a consonant solution, unique on the set of solutions found, that corresponds to the model proposed by G. SHAFER on the basis of this characteristic alone, for a context similar to that of the present Axiom 3 [1]. To give a practical expression to this solution, let us suppose that the $p(s_j/H_i)$ are arranged such that $p(s_j/H_1) \geq p(s_j/H_2) \geq \dots \geq p(s_j/H_I)$. The focal elements are the I subsets of E :

$$A_i = \bigcup_{k \leq i} H_k, \quad i \in [1, I] \quad (B10)$$

and the corresponding masses are given by :

$$m_j(A_i) = K_j * p(s_j/H_i) \quad (B11)$$

$$m_j(A_i) = K_j * \{ p(s_j/H_i) - p(s_j/H_{i+1}) \}, \quad \text{pour } 1 \leq i \leq I-1 \quad (B12)$$

It should nonetheless be pointed out that this last solution does not satisfy Axioms 4 and 5, and that it therefore cannot be retained in the following.

Let us lastly say that, in the ideal case where the distributions $p(s_j/H_i)$ are perfectly representative of the densities actually encountered, a maximum likelihood procedure requires retaining the hypothesis H_i that will maximize $p(s_1, \dots, s_I/H_i)$. Yet since the hypotheses H_i are singletons of the frame of discernment E , and $p(s_1, \dots, s_I/H_i)$ is the product of the $p(s_j/H_i)$ provided by the I independent sources S_j , the plausibility $Pl(H_i)$ obtained after associating the sources S_j is expressed, using (B6), by :

$$Pl(H_i) = K_j * p(s_1, \dots, s_I/H_i), \quad \forall i \in [1, I] \quad (B13)$$

in which the coefficient K_j , independent of H_i , integrates the K_j terms and the inconsistency of the combination. To remain consistent with this particular case, any decision procedure to designate the most realistic hypothesis must, for our problem, exclusively maximize a monotonic increasing function of the plausibility $Pl(H_i)$ alone, obtained after combining the sources S_j .

Axiom 4 : Separability of Hypothesis Evaluations

This axiom consists in considering that each mass function $m_j(\cdot)$ sought is itself the result of a combination between I mass functions $m_{ij}(\cdot)$ ($i \in [1, I]$) :

$$m_j(\cdot) = \bigoplus_i m_{ij}(\cdot) \quad (B14)$$

A mass function $m_{ij}(\cdot)$ also has three focal elements (H_i , $\neg H_i$, and E), whose masses depend only on the value $p(s_j/H_i)$ and the corresponding factor q_{ij} .

Since the hypotheses H_i are the singletons of the frame of discernment E , the plausibility $Pl_j(H_i)$ is proportional to the product on k of the $Pl_{kj}(H_i)$ associated with the $m_{kj}(\cdot)$. After factorization of the product on k of all the $Pl_{kj}(\neg H_k)$, it is finally expressed :

$$Pl_j(H_i) = K_{fj} * \{ m_{ij}(H_i) + m_{ij}(E) \} / \{ m_{ij}(\neg H_i) + m_{ij}(E) \}, \quad i \in [1, I] \quad (B15)$$

in which the factor K_{fj} is independent of the hypothesis H_i concerned.

Holding to constraint (A6) as required by Axiom 3 will then permit the probability $p(s_j/H_i)$ alone to be

associated with mass function $m_{ij}(\cdot)$ alone, for $q_{ij}=1$, only if :

$$\{m_{ij}(H_i) + m_{ij}(E)\} / \{m_{ij}(-H_i) + m_{ij}(E)\} = R_j * p(s_j/H_i) \quad (B16)$$

in which R_j is a normalization constant independent of H_i , whose possible values depend only on the distributions $p(s_j/H_i)$ actually taken into account, as we shall see in the following. In practice, this constant allows us to consider the general framework where the $p(s_j/H_i)$ are known only relatively, i.e. to within a normalization gain.

Expressed parametrically as a function of the level of uncertainty $m_{ij}(E)$, (B16) procures the desired mass function :

$$m_{ij}(H_i) = \{R_j * p(s_j/H_i) - m_{ij}(E)\} / \{1 + R_j * p(s_j/H_i)\} \quad (B17)$$

$$m_{ij}(-H_i) = \{1 - R_j * p(s_j/H_i) * m_{ij}(E)\} / \{1 + R_j * p(s_j/H_i)\} \quad (B18)$$

$$m_{ij}(E) = f[R_j * p(s_j/H_i)] \in [0, R_j * p(s_j/H_i)] \quad (B19)$$

in which f is any function verifying simply (B19).

This condition (B19) is required by the mass idea (included between 0 and 1), which also limits the possible values of R_j as a function of the distributions $p(s_j/H_i)$ used, and does so independently of the measures s_j actually observed :

$$R_j \in [0, (\max_{s_j, i} \{p(s_j/H_i)\})^{-1}] \quad (B20)$$

It is furthermore possible to show that these conditions are sufficient in order for the coefficient K_j in expression (B6), calculated for the combination (B14), to verify the constraint (B7). This can be done simply by showing that the expression for K_j is then an increasing monotonic function of each $m_{ij}(E)$, whose extreme values make it possible to satisfy the interval (B7).

If we introduce the factor q_{ij} into the expressions (B17), (B18), and (B19) in terms of discounting, the $m_{ij}(\cdot)$ are finally given by :

$$m_{ij}(H_i) = q_{ij} * \{R_j * p(s_j/H_i) - A_i\} / \{1 + R_j * p(s_j/H_i)\} \quad (B21)$$

$$m_{ij}(-H_i) = q_{ij} * \{1 - R_j * p(s_j/H_i) * A_i\} / \{1 + R_j * p(s_j/H_i)\} \quad (B22)$$

$$m_{ij}(E) = 1 - q_{ij} + q_{ij} * A_i \quad (B23)$$

in which R_j is still defined by (B20), and A_i by :

$$A_i = f[R_j * p(s_j/H_i)] \in [0, R_j * p(s_j/H_i)] \quad (B24)$$

The general expression of the models $m_j(\cdot)$ that satisfy Axioms 3 and 4 is thus found by (B14) applied to

(B21), (B22), and (B23). An infinite number of solutions thus still fit our problem.

Axiom 5: Consistency With the Probabilistic Association of the Sources

Considering the special structure (B14) of the mass functions $m_j(\cdot)$ complying with Axioms 3 and 4, and the associativity of the orthogonal sum, Axiom 5 will be satisfied for models such that, if the q_{ij} are equal to 1, the mass function $m_i(\cdot)$ defined by :

$$m_i(\cdot) = \bigoplus_j m_{ij}(\cdot) \quad (B25)$$

$$m_{ij}(\cdot) = F[R_j * p(s_j/H_i)] \quad (B26)$$

is identical to the mass set $m'_i(\cdot)$ obtained by direct modeling, using the same function $F(\cdot)$:

$$m'_i(\cdot) = F[\prod_j \{R_j * p(s_j/H_i)\}] \quad (B27)$$

The $m_{ij}(\cdot)$ verifying (B17), (B18), (B19), and (B20), in the combination (B25), yield :

$$m_i(H_i) = (V * X - Y * W) / (V * X + X - Y * W) \quad (B28)$$

$$m_i(-H_i) = (X - Y * W) / (V * X + X - Y * W) \quad (B29)$$

$$m_i(E) = Y * W / (V * X + X - Y * W) \quad (B30)$$

with the definitions :

$$V = \prod_j \{R_j * p(s_j/H_i)\} \quad (B31)$$

$$W = \prod_j \{1 + R_j * p(s_j/H_i)\} \quad (B32)$$

$$X = \prod_j \{1 + m_{ij}(E)\} \quad (B33)$$

$$Y = \prod_j m_{ij}(E) \quad (B34)$$

and the constraints :

$$m_{ij}(E) = f[R_j * p(s_j/H_i)] \in [0, R_j * p(s_j/H_i)] \quad (B35)$$

$$R_j \in [0, (\max_{s_j, i} \{p(s_j/H_i)\})^{-1}] \quad (B36)$$

At the same time, the mass set $m'_i(\cdot)$ is written :

$$m'_i(H_i) = \{V - m'_i(E)\} / \{1 + V\} \quad (B37)$$

$$m'_i(-H_i) = \{1 - V * m'_i(E)\} / \{1 + V\} \quad (B38)$$

$$m'_i(E) = f[\prod_j \{R_j * p(s_j/H_i)\}] \in [0, \prod_j \{R_j * p(s_j/H_i)\}] \quad (B39)$$

in which V is still given by (B31), and the R_j are also constrained by (B36).

We can go about comparing the mass sets $m_i(\cdot)$ and $m'_i(\cdot)$ by letting $m_i(E)=m'_i(E)$ in (B30). Then expressions (B37) and (B38) are equivalent to expressions (B28) and (B29), respectively, which means that under all circumstances $m_i(H_i)=m'_i(H_i)$ and $m_i(\neg H_i)=m'_i(\neg H_i)$. On the other hand, (B35) and (B39) will be equivalent for the same function f , through (B30) still under the constraint $m_i(E)=m'_i(E)$, only for the following two functions f :

$$f(x) = 0, \forall x \quad (B40)$$

$$f(x) = x \quad (B41)$$

After examination of Axiom 5, only two models are left that simultaneously satisfy the three axioms. Both are defined by (B21), (B22), (B23). They differ by the fact that $A_i=0$ for one while $A_i=R_j \cdot p(s_j/H_i)$ for the other, the R_j being constrained by (B20) in both cases.

Summary of the Models Obtained

There are finally only two models, then, that jointly satisfy the three desired axioms. Both meet the decomposition :

$$m_j(\cdot) = \bigoplus_i m_{ij}(\cdot) \quad (B42)$$

Model 1 is particularized by :

$$m_{ij}(H_i) = 0 \quad (B43)$$

$$m_{ij}(\neg H_i) = q_{ij} \cdot \{1 - R_j \cdot p(s_j/H_i)\} \quad (B44)$$

$$m_{ij}(E) = 1 - q_{ij} + q_{ij} \cdot R_j \cdot p(s_j/H_i) \quad (B45)$$

and Model 2 by :

$$m_{ij}(H_i) = q_{ij} \cdot R_j \cdot p(s_j/H_i) / \{1 + R_j \cdot p(s_j/H_i)\} \quad (B46)$$

$$m_{ij}(\neg H_i) = q_{ij} / \{1 + R_j \cdot p(s_j/H_i)\} \quad (B47)$$

$$m_{ij}(E) = 1 - q_{ij} \quad (B48)$$

In both cases, the normalization factor R_j is simply constrained by :

$$R_j \in [0, (\max_{s_j, i} \{p(s_j/H_i)\})^{-1}] \quad (B49)$$

Appendix C : Filter Expression

The filtering, prediction, and statistical gating modules are those of a PDAF that would operate at minimum threshold with $P_d=P_{fa}=1$.

Filtering

The α^m coefficients are given by :

$$\alpha^0 = M \cdot (2 \cdot \pi / \gamma)^{r/2} \cdot (1 - P_g) / C_r \quad (C1)$$

$$\alpha^m = \exp[-0.5 \cdot (x^m - x_k)^T V_k^{-1} (x^m - x_k)] \quad (C2)$$

$$\text{in which : } C_r = \pi^{r/2} / \Gamma(1 + r/2) \quad (C3)$$

x_k and V_k designate the predicted position and its covariance, at time k . r is the common dimension of x_k and x^m . P_g represents the *a priori* probability that the target is in the validation gate, considering the choice of statistical gating threshold γ .

The β^m coefficients are determined from the α^m coefficients by (7.1) to (7.7). The estimated state $X_{k/k}$ and its covariance $P_{k/k}$, which are outputs of the procedure, are then updated at time k by :

$$X_{k/k} = X_{k/k-1} + G_k \cdot z_k \quad (C4)$$

$$P_{k/k} = \beta^0 \cdot P_{k/k-1} + (1 - \beta^0) \cdot (I - G_k \cdot H) \cdot P_{k/k-1} + P_k \quad (C5)$$

in which :

$$P_k = G_k \cdot [(\sum_{m \neq 0} \beta^m \cdot z_k^m \cdot z_k^{mT}) - z_k^* \cdot z_k^{*T}] \cdot G_k^T \quad (C6)$$

$$z_k^m = x^m - x_k \quad (C7)$$

$$z_k = \sum_{m \neq 0} (\beta^m \cdot z_k^m) \quad (C8)$$

$$G_k = P_{k/k-1} \cdot H^T \cdot V_k^{-1} \quad (C9)$$

H is the position observation matrix.

Prediction

The predicted state, $X_{k/k-1}$, and its covariance, $P_{k/k-1}$, used above for updating the filter, are calculated from the state $X_{k-1/k-1}$ and its covariance $P_{k-1/k-1}$ estimated at the time of previous observation $k-1$ by the filtering module :

$$X_{k/k-1} = F \cdot X_{k-1/k-1} \quad (C10)$$

$$P_{k/k-1} = F \cdot P_{k-1/k-1} \cdot F^T + Q \quad (C11)$$

in which F is the state transition matrix from one observation time to the next one, and Q the noise covariance matrix on the state.

The predicted position measurement x_k and its covariance V_k , used by the filtering and gating modules, are then determined by :

$$x_k = H \cdot X_{k/k-1} \quad (C12)$$

$$V_k = H \cdot P_{k/k-1} \cdot H^T + R \quad (C13)$$

with R designating the noise covariance matrix on the position measurement.

Statistical Gating

The cells x^{ln} and x^m to be processed (figure 11) are selected by the tests :

$$(x^{ln-xk})^T V_k^{-1} (x^{ln-xk}) \geq \gamma \quad (C14)$$

$$(x^m-xk)^T V_k^{-1} (x^m-xk) \geq \gamma \quad (C15)$$

Appendix D : Combination Process for Tracking

The elaboration of the combination processing suitable for the tracking problem tackled in section 7 is developed according to the scheme provided in figure 11.

Procedure at the Level of Each Resolution Cell

The first step consists in associating the mass functions $m_{ij}^{ln}(\cdot)$ defined on a same frame of discernment $E_i^{ln} = \{H_i^{ln}, -H_i^{ln}\}$, as regards the various sensors j of a given alignment group l . For each E_i^{ln} , their orthogonal sum directly yields the mass function $m_i^{ln}(\cdot)$ defined by :

$$m_i^{ln}(H_i^{ln}) = \left\{ \prod_{j=1}^J [m_{ij}^{ln}(H_i^{ln}) + m_{ij}^{ln}(E_i^{ln})] - \prod_{j=1}^J m_{ij}^{ln}(E_i^{ln}) \right\} / (1 - K_i^{ln}) \quad (D1)$$

$$m_i^{ln}(-H_i^{ln}) = \left\{ \prod_{j=1}^J [m_{ij}^{ln}(-H_i^{ln}) + m_{ij}^{ln}(E_i^{ln})] - \prod_{j=1}^J m_{ij}^{ln}(E_i^{ln}) \right\} / (1 - K_i^{ln}) \quad (D2)$$

$$m_i^{ln}(E_i^{ln}) = \left\{ \prod_{j=1}^J m_{ij}^{ln}(E_i^{ln}) \right\} / (1 - K_i^{ln}) \quad (D3)$$

in which K_i^{ln} represents the inconsistency of the combination, the expression of which is not necessary for the remaining discussion.

The refinement of E_i^{ln} in the set $E^{ln} = \{H_1^{ln}, \dots, H_I^{ln}\}$ of the identities attached to the cell x^{ln} allows the minimum extension, in the common frame of discernment E^{ln} , of the $m_i^{ln}(\cdot)$ relative to these different identities. The orthogonal sum of the mass functions obtained leads to the mass function $m^{ln}(\cdot)$. Then all we have to do is express the plausibilities of H_I^{ln} and of

$-H_I^{ln}$, which are all that is needed for the rest of the discussion :

$$Pl^{ln}(H_I^{ln}) = (1 - K^{ln})^{-1} * [m_I^{ln}(H_I^{ln}) + m_I^{ln}(E_I^{ln})] * \prod_{i=1}^{I-1} [m_i^{ln}(-H_i^{ln}) + m_i^{ln}(E_i^{ln})] \quad (D4)$$

$$Pl^{ln}(-H_I^{ln}) = (1 - K^{ln})^{-1} * [m_I^{ln}(-H_I^{ln}) + m_I^{ln}(E_I^{ln})] * \prod_{i=1}^{I-1} [m_i^{ln}(-H_i^{ln}) + m_i^{ln}(E_i^{ln})] - \prod_{i=1}^{I-1} m_i^{ln}(-H_i^{ln}) * \prod_{i'=1, i' \neq i}^{I-1} [m_{i'}^{ln}(H_{i'}^{ln}) + m_{i'}^{ln}(E_{i'}^{ln})] \quad (D5)$$

in which K^{ln} is the combination inconsistency, the expression of which is not useful for the following.

Summary at the Level of the Validation Gate

So two cases should be distinguished, depending on whether the x^{ln} resolve the validation gate or not (see introduction of section 7). If they do, then the E^{ln} relative to each x^{ln} cell need only be refined in the set $E^l = E^{l1} \times \dots \times E^{lN}$ of possible identity distributions on the cells in question, and the orthogonal sum of the resulting minimum extensions can be performed. The very special nature of the associated focal elements, each being specific to a distinct component of the cartesian product, allows a relatively simple expression for the only plausibilities we now have to evaluate on the basis of (D4) and (D5). These plausibilities concern the N hypotheses H_I^{ln} of the presence of identity H_I respectively in the cell x^{ln} , to the exclusion of any other cell, and the hypothesis H^0 of the absence of identity H_I in the gate. These hypotheses are in fact specific subsets of E^l , as there exists one and only one target of identity H_I in the gate, according to the axiom adopted to start with (see introduction of section 7). This leads to :

$$Pl^l(H^0) = \prod_{n=1}^N Pl^{ln}(-H_I^{ln}) \quad (D6)$$

$$Pl^l(H_I^{ln}) = Pl^{ln}(H_I^{ln}) * \prod_{\substack{n'=1 \\ n' \neq n}}^N Pl^{ln'}(-H_I^{ln'}) \quad (D7)$$

If the x^{ln} do not resolve the validation gate, an additional prior refinement should be performed from each E^{ln} to a set $E'^{ln} = \{H_1^{ln0}, H_1^{ln1}, \dots, H_I^{ln0}, H_I^{ln1}\}$ to split each hypothesis H_I^{ln} between, on the one hand, a similar hypothesis H_I^{ln1} simply relative to the part of

x^{ln} covering the gate, and, on the other hand, an additional hypothesis H_I^{ln0} relative to the part of x^{ln} outside the gate. The operations conducted in the case where the gate is resolved are then conducted on the modified set $E^l = E^l 1_{x^{ln}} \dots x^{ln} E^l$. However, hypothesis H_I^{ln} is now reduced to the presence of the identity H_I just in that part of cell x^{ln} covering the gate, and excluding any other cell defined in the gate. Hypothesis H^0 , though its definition remains unchanged, also corresponds to a different subset of E^l . The result is the modified expressions :

$$Pl(H^0) = 1 \quad (D8)$$

$$Pl(H^{ln}) = Pl(H_I^{ln}) \quad (D9)$$

The groups l of unaligned sensors are then combined in both cases by refining the E^l in the common set $E^F = E^1 x^{ln} \dots x^{ln} E^L$, and performing the orthogonal sum of the associated minimal extensions. This step reflects the fact that we are interested in the intersections x^m of the cells x^{ln} (see section 7.1). As each of the associated focal elements here is still specific to a distinct component of the cartesian product, the plausibilities of H^0 and of the hypotheses H^m of presence of identity H_I in cell x^m , to the exclusion of any other cell, are expressed for the resulting mass function $m^F(\cdot)$ by :

$$Pl^F(H^0) = \prod_{l=1}^L Pl^l(H^0) \quad (D10)$$

$$Pl^F(H^m) = \prod_{l=1}^L Pl^l(H^{ln}) \quad (D11)$$

$$x^m \subset x^{ln}$$

Determination of Weightings β^m and Practical Expression for the Filter

The last step consists in combining this result with the *a priori* probabilities α^m , which constitute a Bayesian mass function on $E^0 = \{H^0, H^1, \dots, H^M\}$. As E^0 is a partition of a subset of E^F , conditioning and coarsening $m^F(\cdot)$ from E^F to E^0 makes it possible to take the orthogonal sum with the set of the α^m . The resulting mass function, which is Bayesian over E^0 , is directly the set of probabilities β^m we are looking for :

$$\beta^0 = \alpha^0 * Pl^F(H^0) / \{ \alpha^0 * Pl^F(H^0) + \sum_{m=1}^M \alpha^m * Pl^F(H^m) \} \quad (D12)$$

$$\beta^m = \alpha^m * Pl^F(H^m) / \{ \alpha^0 * Pl^F(H^0) + \sum_{m'=1}^M \alpha^{m'} * Pl^F(H^{m'}) \} \quad (D13)$$

Expressions (D1) to (D13) can then be summarized by :

$$\beta^0 = \alpha^0 / \{ \alpha^0 + \sum_{m=1}^M \alpha^m * Q^m \} \quad (D14)$$

$$\beta^m = \alpha^m * Q^m / \{ \alpha^0 + \sum_{m'=1}^M \alpha^{m'} * Q^{m'} \} \quad (D15)$$

in which :

$$Q^m = Pl^F(H^m) / Pl^F(H^0) = \prod_{l=1}^L Q^{ln} \quad (D16)$$

$$x^m \subset x^{ln}$$

with, for the sensor groups l that resolve the gate :

$$Q^{ln} = \prod_{j=1}^J A_{ij}^{ln} / \{ 1 - \prod_{i=1}^{I-1} (1 - \prod_{j=1}^J B_{ij}^{ln}) + \sum_{i=1}^{I-1} (\prod_{j=1}^J A_{ij}^{ln} - \prod_{j=1}^J B_{ij}^{ln}) \} \quad (D17)$$

and, for the sensor groups l that do not resolve the gate :

$$Q^{ln} = \prod_{j=1}^J A_{ij}^{ln} / \{ 1 - \prod_{i=1}^I (1 - \prod_{j=1}^J B_{ij}^{ln}) + \sum_{i=1}^I (\prod_{j=1}^J A_{ij}^{ln} - \prod_{j=1}^J B_{ij}^{ln}) \} \quad (D18)$$

In both cases, the coefficients A_{ij}^{ln} and B_{ij}^{ln} represent, respectively, the expressions :

$$A_{ij}^{ln} = \{ m_{ij}^{ln}(H_I^{ln}) + m_{ij}^{ln}(E_I^{ln}) \} / \{ m_{ij}^{ln}(-H_I^{ln}) + m_{ij}^{ln}(E_I^{ln}) \} \quad (D19)$$

$$B_{ij}^{ln} = m_{ij}^{ln}(E_I^{ln}) / \{ m_{ij}^{ln}(-H_I^{ln}) + m_{ij}^{ln}(E_I^{ln}) \} \quad (D20)$$

Partial Differential Equations for Multiscale Image and Video Processing

Gary Hewer
Naval Air Warfare Center
China Lake, CA 93555
and

Charles Kenney *
ECE Department
University of California
Santa Barbara, CA 93106

July 16, 1999

Abstract

Three PDE image processing methods are discussed: Mumford-Shah variational methods, peer group averaging (PGA), and Osher-Rudin shock filtering. Each of these methods is used in segmenting images into homogeneous regions separated by distinct boundaries; reducing the image to regions and boundaries extracts the image structure in a way that can be interpreted automatically by parsing algorithms.

Mumford-Shah algorithms approach the problem of segmentation as one of approximation. The approximation to the image is represented by the homogenized regions and their boundaries. The best approximation is found by minimizing an objective function that controls 1) degree of approximation, 2) smoothness within regions and 3) extent of region boundaries. These three objectives are controlled by weight parameters; choosing these parameters correctly is a problem of major concern for Mumford-Shah algorithms. Once the parameters are selected the objective function can be minimized via an energy descent method resulting in a nonlinear PDE with the original image as initial data. Evolution under the PDE produces the desired approximation and image segmentation. Recent work by Hewer et al. has implemented this procedure in a manner that minimizes the parameter selection problem and greatly reduces the number of descent steps needed for an acceptable approximation. This reduction is due in part to the use of PGA as a preprocessing step: the initial image data is replaced by the PGA-filtered image. Since PGA produces results that are

near the desired Mumford-Shah approximation only a few descent steps are required.

Peer group averaging is a discrete approximation method that starts with the initial image and then makes processing decisions based on the local peer group. This peer group is determined by nearness in intensity value: the gray level of the central pixel of a local window is compared with the other pixels in the window and the closest ones form the peer group. The average over the peer group is then used to replace the central pixel intensity value. This procedure converges quickly; usually only two or three iterations are needed. In this method the parameters consist only of the window size and the number of pixels in the peer group. These parameters are easily selected for enhancing specific targets.

PGA is closely related to the shock filtering method of Osher and Rudin. In shock filtering image information moves outward from the centers of regions. This outward motion forms standing shock fronts at the boundaries of regions. The convective PDE describing this evolution arises naturally in many situations including flame front propagation and crystal growth. Shock filtering preserves edge location (unlike many noise reduction methods such as Gaussian smoothing) and at the same time maintains the total variation of the image. This means that it removes noise while enhancing the contrast across edges. This effect is desirable in ATR applications. PGA and shock filtering have been shown to be equivalent for 1d signal problems.

Applications are presented to illustrate the Mumford-Shah, shock filtering and PGA image processing methods.

*This research was supported by the Office of Naval Research under ONR Grant Number N00014-96-1-0456.

1 Introduction

In the last decade, a new image processing tool has been developed that uses partial differential equations to generate scale space decompositions of an image. Such decompositions are analogous to the multiresolution decompositions provided by wavelets or Fourier transforms except that the scale parameter for PDE processing is the time evolution under the PDE with the original image as the initial data.

Variational scale space image decompositions are described in Section 2. This approach utilizes an objective function $E = E(g, u, B)$ that depends on the original image g as well as an approximation u and a boundary function B . Typically E contains a penalty term that measures the difference between g and u , another penalty term for the nonsmoothness in u and also a penalty term for the length of the boundaries of the regions in the image. This latter term is needed to control the number of region components in the final segmentation: too many components and the result is not useful.

A variety of methods have been developed to implement this approach. These include region merging schemes used by Koepfler and others [16], [17], applying homotopy type methods to the objective function in order to guarantee convergence of descent methods to global minimizers such as the GNC approach of Blake and Zisserman [4], and using the Euler-Lagrange PDE associated with the objective function with the boundary B interpreted as a continuous function rather than a binary process. For general references to variational methods and PDEs related to image processing see [21] and [29].

Each of these approaches has advantages and disadvantages. For example, region merging generally produces excellent results and is easily adapted to a multichannel form that can accept multiresolution or multispectral image data as input; however region merging is computationally intensive and may not be appropriate for real-time applications. Time considerations also place limitations on the number of iterations that can be used in steepest descent procedures for other variational methods. This means that we must usually forego the luxury of finding the global minimizer of the objective functional and instead seek an approximation that is acceptable rather than optimal.

One way of handling this problem is to modify the image prior to applying the PDE. For example if we desire a very uniform approximation with few regions then Gaussian smoothing provides a fast prefilter. Unfortunately this type of smoothing degrades edges and can shift their position; this means that the prefiltered image is somewhat removed from the variational minimizer and thus the number of PDE

descent steps needed to reach the minimizer may not be reduced sufficiently for speedy computation.

What is needed then is a prefiltering technique that can quickly smooth interior regions without degrading or moving edges. This brings us to peer group averaging (PGA). In this method the pixel intensities are adjusted based on local peer groups so that edges are respected. Typically this method converges very quickly, usually within 2 or 3 iterations. As such it provides an excellent prefilter for variational methods and acts as a starting point for a variety of other applications. PGA is described in Section 3.

Historically PGA was preceded by a closely related method of Osher and Rudin called shock filtering. This method uses a nonlinear convection PDE to propagate information from the interiors of regions in a way that smooths the region. At the boundaries of regions a standing shock forms; this leads to contrast enhancement at the edges and preserves the total variation of the original image. This method and its connection to PGA are described in Section 4.

2 Variational Approximation and Boundary Description

A general variational framework for image segmentation and approximation has been developed by Hewer et al. [14] that simplifies and systematizes approaches that had previously been considered separately, especially those with Mumford-Shah objective functionals [22], [23], [24] and those considered by Geman and others [10], [11], [12].

To set the stage, suppose that we are given a blurred image g over a domain Ω :

$$g = Au_0 + \eta \quad (1)$$

where A is the blurring operator, u_0 is the unblurred image and η is the noise. One approach to segmenting and approximating such an image consists of finding an approximation u and a boundary set K that minimizes an objective functional of the form

$$\begin{aligned} \tilde{E}(u, K) = & w_1 \int_{\Omega \setminus K} (Au - g)^2 + w_2 \int_{\Omega \setminus K} \nabla u \cdot \nabla u \\ & + w_3 \int_K d\sigma \end{aligned} \quad (2)$$

where the last integral term corresponds to the length of the boundary. The scalars w_1, w_2 and w_3 are weighting factors that determine respectively how closely Au approximates g , the smoothness of u and the extent of the boundary. Without loss of generality we may assume that $w_3 = 1$. Functionals of this

type are often referred to as a Mumford-Shah functionals. See [21] p.24, [22], [23] and [24] for details.

Unfortunately numerical procedures for minimizing the Mumford-Shah functional encounter bookkeeping problems associated with tracking regions and their boundaries. These problems can be traced to the binary nature of the boundary description as embodied in the boundary characteristic function χ , which takes on the value 1 on the boundary K and 0 elsewhere. Binary descriptions of boundaries may be appropriate in some special cases but for most problems the transitions between regions can occur over several pixels rather than abruptly. Moreover the mathematical view of the boundary as the differential of a region (hence the notation ∂R for the boundary of a region R) underscores the inherent sensitivity of the boundary description process; this is entirely analogous to the sensitivity of derivatives with respect to noise.

For these reasons, it often is appropriate to specify boundaries with a function B taking continuous values between 0 and 1. Such a function might be viewed as a probability boundary description but we do not explore that issue. Instead our main concerns are utility and ease of numerical computation.

To accommodate a continuous boundary function B , the Mumford-Shah functional could be recast as

$$E(u, B) = w_1 \int_{\Omega} (Au - g)^2 (1 - B)^2 + w_2 \int_{\Omega} \nabla u \cdot \nabla u (1 - B)^2 + \int_{\Omega} B^2 \quad (3)$$

where w_1 and w_2 are scalar weights. Here we have replaced the integrals over $\Omega \setminus K$ by integrals over Ω with integrands multiplied by $(1 - B)^2$, the idea being that since $B \approx 1$ is near K , the integration of terms times $(1 - B)^2$ over K is nearly 0. Similarly the boundary length integral has been replaced by the integral of B^2 .

There is a significant amount of related work in image processing and vision. Early work in this area dealt with scale space decompositions induced by Gaussian smoothing operators and the motion of edges (as identified with zero-crossings of the Laplacian) in scale space. See [18], [19], [33], [40], and [37].

Identifying spatial discontinuities is helpful in many applications such as segmentation, optical flow, stereo, and image reconstruction. The concept of a "line process" is useful in studying these problems as one of regularization. The binary line process was introduced by Geman and Geman [10] where the authors considered simulated annealing based algorithms for achieving the global optimization. Since then several modifications of the original scheme have

been suggested. Blake and Zisserman [4] formulated the same problem as minimizing an objective functional which enforces smoothness while eliminating the binary line process. See also Geiger and Gersosi [9], Geman and Reynolds [12], and Rangarajan and Chellappa [30]. Some of these recent works involve analog or continuous line processes. The connections between the line process approach to regularization and outlier processes in robust statistics is explored by Black and Rangarajan [3].

Common to all these algorithms is an objective functional that:

- (a) enforces closeness to the original data by including terms such as $(u - g)^2$ or $(Au - g)^2$
- (b) promotes local smoothness away from edges by including terms depending on $\|\nabla u\|$
- (c) limits the extent of the boundary.

For example, Richardson [31] and Richardson and Mitter [32] consider minimizing functionals of the form

$$E_c(u, v) = \int_{\Omega} \beta(u - g)^2 + \Phi(v) \|\nabla u\|^2 + \alpha \left(c \Psi(v) \|\nabla v\|^2 + \frac{(1 - v)^2}{4c} \right) \quad (4)$$

where α, β and c are weighting factors and v is a continuous function describing the boundary. Ambrosio and Tortorelli [1], [2] have shown that, for $\Phi(v) = v^2$ and $\Psi(v) = 1$, this functional "Γ-converges" as $c \rightarrow 0$ to the following form of the Mumford-Shah functional:

$$E(u, v) = \int_{\Omega} \beta(u - g)^2 + \int_{\Omega \setminus K} \|\nabla u\|^2 + \alpha |K| \quad (5)$$

where $|K|$ is the length of the boundary K .

In a similar vein, Shah [36] proposed minimizing a pair of functionals dependent on u and v : Given u find v minimizing

$$V_u(v) = \int_{\Omega} \alpha(1 - v)^2 \|\nabla u\| + \frac{\rho}{2} \|\nabla v\|^2 + \frac{v^2}{2\rho} \quad (6)$$

where α and ρ are weighting parameters. Given v find u minimizing

$$U_v(u) = \int_{\Omega} \|\nabla u\|^2 + \frac{(u - g)^2}{v^2 \sigma^2} \quad (7)$$

where σ is a weighting parameter.

The idea of the second functional is that the boundary function v is approximately 0 inside regions where we want $u - g$ to be small. Hence the

division of $(u - g)^2$ by v^2 can be interpreted as a local weighting that enforces close approximation of g inside regions. Applying a steepest decent minimization procedure to these functionals yields a pair of coupled diffusion PDEs for u and v . This is also the case for the functionals studied by Richardson and Mitter.

However, as noted by Proesmans et al. [29], Shah's approach leads to blurring of the edges; this can be partially offset by working with a modified objective functional but some blurring still remains.

This blurring effect appears to be induced in part by the inclusion of the boundary gradient term $\|\nabla v\|$ in the objective functional, since this results in a diffusion PDE for v . Inclusion of the boundary gradient term also has the effect of "masking" the boundary. That is, for a given approximation function u the optimal boundary function v is the solution of a nonlinear elliptic PDE and cannot be given explicitly.

In contrast, the objective functionals of the Mumford-Shah type (4) as well as objective functionals of the type considered by Geman and Reynolds [12], which extend the work of Geman et al. [11], do not include a boundary gradient term.

2.1 Reducible Objective Functionals

Consider the following generalized form of the Mumford-Shah functional (4)

$$E(u, B) = \int_{\Omega} r (1 - B)^2 + B^2 \quad (8)$$

where the residual term r depends on $Au - g$ as well as ∇u . For our purposes we have found the following form of r to be most useful

$$r = w_1(Au - g)^2 + w_2\|\nabla u\| \quad (9)$$

but more general forms of r are also considered below. Functionals of this type have the big advantage that the optimal boundary function B can be found explicitly for any nonnegative residual function r : independent of the form of r we show that, for a given function u , the function B that minimizes $E(u, B)$ is given by

$$B = \frac{r}{1 + r}. \quad (10)$$

We denote this optimal boundary function by $B = B(u)$. This allows us to eliminate B from the objective functional and (after some simple algebra) we are led to the equivalent problem of minimizing the functional $E(u) \equiv E(u, B(u))$ given by

$$E(u) = \int_{\Omega} \frac{r}{1 + r}. \quad (11)$$

It is interesting that this reduced functional is equal to the L_1 norm of the optimal boundary function

B ; that is minimization of the reduced functional is really the same as minimizing the L_1 norm of B subject to $B = r/(1 + r)$.

The following lemma shows that there is a unique boundary function that minimizes $E(u, B)$.

Lemma 1 Let $r = r(u, g, \nabla u)$ be nonnegative. For fixed g and u , the objective functional defined by

$$E(u, B) = \int_{\Omega} r (1 - B)^2 + B^2 \quad (12)$$

is minimized by setting $B = r/(1 + r)$. Moreover, for any B

$$E(u, B) \geq \int_{\Omega} \frac{r}{1 + r} \quad (13)$$

with equality only for $B = r/(1 + r)$.

Proof: See [14].

2.2 Numerical Implementation

Once the form of the variational functional has been selected, the nontrivial problem of finding the minimizing approximation u has to be addressed. Typically the desired approximation is an equilibrium solution of a nonlinear diffusion PDE with certain boundary conditions. To illustrate, suppose that we wanted to minimize a functional of the form

$$E(g, u) = \int_{\Omega} (u - g)^2 + \nabla u \cdot \nabla u \quad (14)$$

where g is the given image and u is an approximation of g .

The minimizing approximation u for this functional satisfies the elliptic equilibrium PDE

$$\begin{aligned} \Delta u &= u - g \\ \partial u / \partial n &= 0 \quad \text{on } \partial \Omega \end{aligned}$$

where Δu is the Laplacian of u and $\partial u / \partial n$ denotes the normal derivative on the boundary $\partial \Omega$.

Numerically we can either solve for the equilibrium solution directly or follow u as a function of t from an initial approximation, such as $u_0 = g$, by integrating the diffusion PDE

$$u_t = g - u + \Delta u \quad (15)$$

subject to the Neumann boundary condition $\partial u / \partial n = 0$ on $\partial \Omega$. Starting from the initial condition u_0 the image u evolves as $t \rightarrow \infty$ toward the equilibrium solution.

The numerical results for this paper were obtained using the mixed norm objective functional

$$E = \int_{\Omega} (w_1(Au - g)^2 + w_2\|\nabla u\|) (1 - B)^2 + B^2.$$

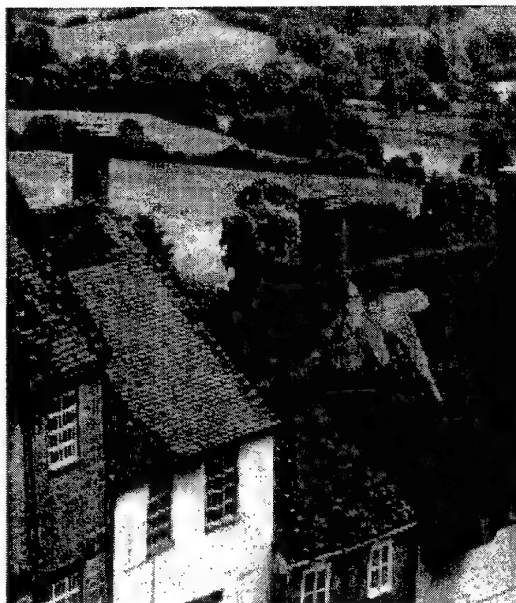


Figure 1: Goldhill image: original (left) and after variational processing (right).

We used the 1-norm for the smoothness term ∇u since this produced sharper edges in the approximation u than the 2-norm. Note that to avoid discontinuous derivatives at $\nabla u = 0$ we use the modified smoothness term $(\nabla u \cdot \nabla u + \delta)^{1/2}$ instead of $\|\nabla u\|$. With this modification, the Euler-Lagrange descent method for this objective functional yields the following PDE. For a given approximation u of g , define the residual

$$r = w_1(u - g)^2 + w_2(\nabla u \cdot \nabla u + \delta)^{1/2} \quad (16)$$

then the descent PDE for u is given by

$$u_t = \frac{2w_1(g - u)}{(1 + r)^2} + w_2 \nabla \cdot \left(\frac{\nabla u}{(1 + r)^2} (\nabla u \cdot \nabla u + \delta)^{-1/2} \right) \quad (17)$$

subject to the Neumann boundary condition.

Euler's method was used to integrate the descent PDE and we halted the integration when the decrease in the value of the objective functional became less than a user supplied tolerance. Typically we obtained good results by stopping when the decrease in the objective functional from one Euler step to the next was less than 1 percent of the current value of the objective functional. This PDE descent procedure may lead to a local minimum for the objective functional. Other methods, such as simulated annealing can be used to find a global minimum for

the objective functional with high probability, but the intensive computational costs can result in unacceptably long processing times. See [14] for details.

Example 1: Figure 1 (left) shows a detail from a standard image entitled "Goldhill". Applying the variational boundary method above using (16) and (18) produces the results in Figure 1 (right). Note that the variational approach has suppressed the small details in the image such as the sheep in the background and the texture of the roof in the foreground. The amount of smoothing and detail suppression is controlled by the weight parameters w_1 and w_2 in (16). Figure 2 shows the associated boundary map.

To avoid the need to take many (usually hundreds) of descent steps in minimizing the objective functional it is helpful to prefilter the image. The next section describes a prefilter that we have used [13] successfully.

3 Peer Group Averaging

Peer group averaging is a fast image processing scheme that enhances objects of a given diameter and area. The basic idea consists of two steps: to enhance objects with n or more pixels 1) identify a peer group of size n for each pixel 2) process the pixel value based on the characteristics of the peer group. There are many ways to select the peer group for a given pixel. For example, see the earlier work by



Figure 2: Goldhill image variational boundary map.

Yaroslavsky [38] presenting an abstract formulation of the group idea. In general, peer group members should share common values. For a single image, the peer group may be nearby pixels with similar intensity values. For a sequence of images used in determining optical flow fields, the peer group can be nearby pixels (in time and space) with similar intensity values and similar velocity values. In another context, texture values may be assigned to each pixel and then the peer group determined by nearness in texture space.

In this paper we discuss peer groups based on intensity nearness. For a given image g , select a window diameter d and a peer group number n . The selection of d and n should correspond to the size of the objects that are to be enhanced. The peer group for a pixel is selected from the window centered at the pixel and consists of the n pixels whose intensity values are closest to the center value. Let u be the average over the peer group. If we let A_k be the averaging operator at step k we can represent the PGA iteration as $u_{k+1} = A_k u_k$ where $u_0 = g$. The PGA iteration is nonlinear because of the peer group selection.

Convergence of the PGA iteration is considered by Deng et al. [6] who show that the PGA iterates converge to an image that is constant on the interior regions of the image (the 'irreducible' subsets of Theorem 2 in [6]). A comparison of PGA with median filtering, the shock filtering method of Osher and Rudin [26] and morphological filtering is given in [6].

One of the main features of PGA is that it is a discrete method designed for images rather than a continuous method such as PDE or variational methods that are subsequently adapted to discrete images. This correspondence between the motivating derivation and the final application means that the parameters of PGA are closely aligned to the image characteristics that we want to enhance. This is dealt with below where we discuss parameter selection.

3.1 Properties of PGA and Parameter Selection

The most immediate property of PGA processing is the invariance of objects for properly selected parameters. That is, if a group of n pixels all have the same intensity value and the maximum distance between pixels is equal to r , then by setting the peer group number set equal to n and the window diameter equal to $d = 2r$, the common intensity value of this group of pixels is preserved under PGA processing. This property is stable with respect to noise in the following sense. If the intensity values of the object are perturbed by noise that is small enough in magnitude so that the membership in the peer group of the object is not changed, then under PGA processing the pixel values of the object converge in one iteration to their collective average. This average value is equal to the true intensity value of the object plus the average of the noise over the object. Suppose that the noise is independently and identically distributed over the pixels with mean 0 and standard deviation σ . Then the mean of intensity value assigned to the object under PGA processing is the same as the true intensity value, with standard deviation σ/\sqrt{n} where n is the number of pixels in the object. From this we see that PGA processing is very effective at damping out noise even for objects with only a small number of pixels.

Edge enhancement algorithms have to balance conflicting demands. On the one hand, edges of important features should be strengthened without changing their location. At the same time, we want to smooth region interiors and reduce undesirable edges associated with clutter and noise.

The extent to which these opposing goals are met is determined by the choice of the algorithm parameters. For PGA there are two parameters: window diameter d and peer group number n . The conflict between preserving edges and smoothing unwanted detail is reflected in the following observations:

1) If n is larger than the number of pixels in an object O then O will be merged with a larger region of size at least n . In this case edges associated with O may be lost or relocated. Thus, to preserve edges in an object O , use $n \leq n(O)$ where $n(O)$ is the number of pixels in O .

- 2) To preserve straight lines of pixel width w , use $n \leq wd$. This follows from identifying the line as an object O and noting that O has at most wd pixels in a window of diameter d .
- 3) To preserve corners of interior angle at least $\pi/2$, identify O with a square corner with point at the center of the window. Use $n \leq d^2/4$.
- 4) For 1d signals, maximal smoothing without edge loss for an object of size $n(O)$ is obtained by setting the window diameter equal to $2n(O) - 1$ and the peer group number equal to $n(O)$. If the window diameter is reduced below $2n(O) - 1$ while $n = n(O)$, then pixels just outside the object will average over some of the object pixels and edge blurring will occur. Taking $n = n(O)$ and $d = 2n(O) - 1$ allows each pixel to select a peer group entirely to the left or entirely to the right (including the pixel itself) thus preserving the edges.
- 5) As the ratio of the peer group number to the window area increases, the PGA approximation becomes smoother.

3.2 Analytic Results on Parameter Selection

The following lemma is useful in analyzing the problem of selecting the peer group number.

Lemma 3 Let R be a region with $n(R)$ pixels. If S_1 and S_2 are subsets of R with both $n(S_1)$ and $n(S_2)$ larger than $n(R)/2$, then S_1 and S_2 have a nonempty intersection.

Proof: See [6].

This lemma is related to the idea of local connectivity. We say that a region R in an image g is *locally connected* for the peer group parameters d and n if the peer group for any pixel i in R has nonempty intersections with the peer groups of the immediate neighboring pixels for each iteration of the PGA algorithm for g . If a region R is locally connected then there are no sharp edges in R since for any two neighboring pixels the common portions of their peer groups ensure that their final values under PGA are related. The next two lemmas discuss conditions on the PGA parameters that lead to local connectivity for all or part of the image.

Lemma 4 If the peer group number n is large relative to the window diameter d , then the entire image is locally connected under PGA. This occurs if $n > (d+1)/2$ for 1d signals and if $n > d(d+1)/2$ for 2d images where we assume that d is odd.

Proof: See [6].

From the above it is clear that for a fixed window size, as the peer group number increases so does the

smoothing. The next lemma discusses how the peer group number affects smoothing within regions. The idea here is to recover objects under PGA approximation in the sense that all the pixel values of the object are locally connected to each other. However we also want to avoid connections with pixels outside the object. If $n > n(O)$ where $n(O)$ is the number of pixels in the object O , then the peer groups for pixels in O will be forced to include pixels outside of O , resulting in edge blurring. On the other hand, if the peer group number is too small, then slight variations within O can lead to O breaking up into several smaller unconnected regions under PGA.

To analyze this problem we will make the simplifying assumption that the object O is *well-separated* from the rest of the image in the sense that for any peer group number $n \leq n(O)$ and any pixel $i \in O$, the peer group $P(i)$ is a subset of O .

Lemma 5 Assume that O is an object in an image g that is well-separated from the rest of g . If the peer group number n satisfies $n(O)/2 < n \leq n(O)$ then O is locally connected under PGA.

Proof: See [6].

3.3 Automatic Parameter Selection

Although the preceding observations make it possible to predict in a general way how the peer group size affects the smoothing under PGA, it is still the case that in most images we want to vary the peer group size from point to point in order to enhance some features and smooth others. For example, if we use a 3x3 window then a peer group of size 6 preserves straight edges but not corners. If we lower the peer group number to size 4 then corners are also preserved but we don't achieve the smoothing that we see with $n = 6$.

To get around this problem Deng et al. [5] introduced the idea of using the Fisher discriminant to select the peer group for each pixel. That is for a particular pixel let g_1, g_2, \dots, g_m be the intensity values over the window with g_c the intensity of the central pixel. Form the intensity differences $d_i = |g_i - g_c|$. Use the Fisher discriminant to separate these differences into 2 groups. That is maximize the objective functional

$$F(k) = \frac{|a_1 - a_2|}{v_1 + v_2}$$

over the peer group number k , where a_1 and v_1 are the average and variance over the first group and a_2, v_2 are the average and variance of the second group.

This procedure produces excellent results with only a slight increase in processing time to minimize the Fisher discriminant.



Figure 3: Goldhill image: Fisher-PGA (left) and variational processing (right).

3.4 PGA as a Preprocessor for Variational Approximation

In general, one selects the original image g as initial data for the descent PDE associated with the variational objective functional. However, this can lead to the need to take many (hundreds) of descent steps in order to achieve minimization. To avoid this problem we used PGA as a preprocessing step to generate an initial image for the descent PDE; for details see [13].

Example 2: We applied PGA to the Goldhill image using automatic parameter selection via the Fisher discriminant; see Figure 3 (left). Note that the results are almost indistinguishable from the variational processing in Figure 3 (right). Since the PGA processing is much faster than the variational method we obtain considerable computational savings by using PGA alone or as a preprocessor for the variational method.

3.5 Multiscale PGA

One problem with PGA is the limitation to small windows for computational speed. In particular it would be nice to be able to obtain uniform smoothing over large regions without having to use large windows and peer groups. To achieve this a multiscale PGA procedure similar in spirit to multigrid methods for solving large systems of linear equations has been developed [6]. The basic idea is to work on several levels by defining windows with skips be-

tween pixels. At the first level is the usual window with a distance of 1 between pixels; the next level has a distance of 2 between pixels etc. Alternating the PGA iteration between levels results in speeding the passage of intensity information within regions. Fortunately there is a simple way to implement this procedure. For example to do a PGA iteration with a distance of 2 between pixels in each window, one simply has to subsample the image skipping every other pixel and then run regular PGA on the subsampled image. Subsampling this way transforms a large image into 4 smaller images; after running PGA on each of the smaller images they are then recombined into a larger image. In this way we may do one iteration of PGA on the large image followed by one iteration on the smaller images and repeat until the process converges. Convergence is usually quick (3 to 5 iterations). Further subsampling of the small images can be done if desired.

Example 3: Figure 4 (left) shows a detail from a satellite image of an agricultural area. This image is heavily contaminated by speckle and background clutter. Applying multiscale PGA eliminates the speckle in just 3 iterations as see in Figure 4 (right).

4 Shock Filtering

In shock filtering [34] [26],[27], intensity values from the interior of regions move outward towards the region edges along gradient lines. The convexity



Figure 4: Agricultural image: original (left) and after multiscale PGA (right).

of the intensity along the gradient direction determines the motion direction along the gradient and this direction assignment means that when two regions meet at an edge the image intensity will experience a jump. Thus the edges in the image correspond to stationary shock fronts.

In shock filtering the maximum values of the image intensity and the minimum values move outward from the interior of their regions to meet at the boundaries. This means that the contrast at the edges is maximized. This also means that shock filtering preserves the total variation of the original image.

Shock filtering smooths in the sense that each region assumes a constant value. However, shock filtering does not remove isolated noise such as salt-and-pepper noise, as discussed by Osher and Rudin in [26].

In its simplest form for 1d signals, shock filtering uses the original signal g as initial data for a nonlinear convection equation:

$$u_t = -\text{sgn}(u_{xx}) u_x$$

with $u(x, 0) = g(x)$. In this formulation we must be careful to form derivative approximations from the appropriate direction. Thus if intensity information is to move from right to left, then we want u_x to represent the righthand derivative and we use a forward difference to approximate u_x . Similarly we use

a backward difference if we want intensity information to move from left to right.

Consider a simple Euler update scheme for the shock filter equation: let h be the time step and set $u_i^{new} = u_i + hu_t$. If u is monotone increasing at i and $u_{xx} < 0$ in the sense that $u_{i+1} - 2u_i + u_{i-1} < 0$ then the choice $h = 1/2$ leads to $u_i^{new} = (u_i + u_{i+1})/2$. This is the same result we would get with PGA for a peer group of size $n = 2$ because the convexity condition $u_{i+1} - 2u_i + u_{i-1} < 0$ implies that $|u_{i+1} - u_i| < |u_i - u_{i-1}|$. Similarly, if $u_{xx} > 0$ the choice $h = 1/2$ in the shock filter Euler update leads to the same result as the PGA update: $u_i^{new} = (u_{i-1} + u_i)/2$.

This intersection of shock filtering and PGA for particular parameter choices means that results for one method apply immediately to the other. For example, PGA with $n = 2$ for signals is total variation preserving because the same is true for shock filtering. However, the two methods are not the same for other choices of parameters. In particular PGA with larger peer group sizes automatically incorporates smoothing over the peer group and is able to handle problems such as the isolated intensity spikes of salt and pepper noise.

Example 4: This is a 1d signal example consisting of two steps of different heights and widths together with Gaussian noise (see Figure 5a). Figure 5b shows the exact signal with noise added. After using shock filtering we were able to reconstruct

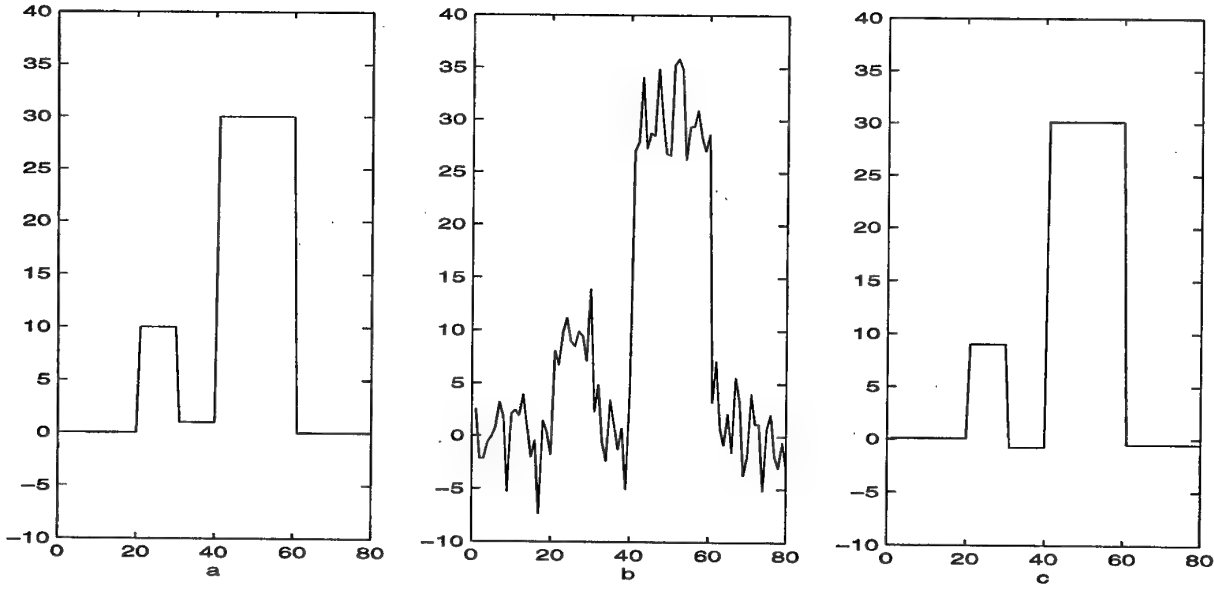


Figure 5: Exact step signal (a), noisy signal (b), shock filter reconstruction (c)

the signal almost exactly (Figure 5c). We also note that a similar example has been studied by Oman [25] using a variety of approximation methods including Sobolev H^1 reconstruction, total variation approximation, low pass Fourier reconstruction, and wavelet methods (in which denoising in the manner of Donoho and Johnstone [7] was used for Harr and Daubechies wavelets). The shock filtering results are superior to (or approximately the same in the case of the total variation method) the results reported by Oman.

The correspondence between PGA and shock filtering is helpful in analyzing the stability of an accelerated version of PGA.

4.1 Shock Filtering and Accelerated PGA

In analogy with the SOR method of solving large systems of linear equations we can accelerate the iterative PGA method using

$$u^{k+1} = \omega \tilde{u}^k + (1 - \omega)u^k$$

where \tilde{u}^k denotes one regular PGA iteration applied to u^k and ω is a scalar between 0 and 2. The stability of this method is rather easily analyzed in the one dimensional case because of the correspondence with shock filtering.

For a convection equation of the form

$$u_t = cu_x \quad (18)$$

the Courant stability condition (c.f. Ferziger [8] p.237) relates the speed of transmission c to the ratio of the spatial step size dx and the temporal step size dt for Eulers method:

$$|c| \leq \frac{dx}{dt}. \quad (19)$$

This can be interpreted as follows: $|c|$ is the speed with which information moves in the solution because the solution to the convection equation has the form $u(x, t) = u(x + ct)$. On the other hand, the ratio dx/dt is the speed with which information moves in the Euler approximation. The Courant condition thus requires that for Euler's method to be stable, information in the approximation must move at least as fast as it does in the true solution. Applying this to the shock filtering equation with $dx = 1$ (i.e., one nodal distance) and $|c| = 1$ we have the stability condition

$$0 \leq dt \leq 1. \quad (20)$$

We can connect the shock filtering stability condition with the accelerated PGA method as follows. Under the same assumptions as for the shock filtering case, i.e., u is montone increasing at i and $u_{xx} < 0$ in the sense that $u_{i+1} - 2u_i + u_{i-1} < 0$ then $|u_{i+1} - u_i| < |u_i - u_{i-1}|$. This means that the peer group for $n = 2$ consists of the pixels i and $i + 1$. This gives the peer group average as

$$a_i = \frac{u_{i+1} + u_i}{2}. \quad (21)$$

Substituting this in the accelerated PGA scheme gives

$$\begin{aligned} u_i^{accel} &= \omega a_i + (1 - \omega)u_i \\ &= \omega \left(\frac{u_{i+1} + u_i}{2} \right) + (1 - \omega)u_i \\ &= \frac{\omega}{2}u_{i+1} + \left(1 - \frac{\omega}{2}\right)u_i. \end{aligned}$$

In the analysis of the shock filtering we assumed that $u_{xx} < 0$. The case for $u_{xx} > 0$ can be handled in the same way except that the averaging is to the left instead of the right. In any case we have the result that the accelerated PGA method with $n = 2$ is stable for 1d signals if $0 \leq \omega \leq 2$ since this range is equivalent to the stability condition $0 \leq dt \leq 1$ for the shock filtering problem.

Accelerated PGA for images takes a particularly simple form for $\omega = 2$ and $n = 2$. Indeed if $\omega = 2$ and $n = 2$, then the update scheme for pixel (i, j) consists of simply selecting the pixel value from the window that is closest in intensity to the central pixel's intensity:

$$u_{ij}^{accel} = u_{i'j'} \quad (22)$$

where $(i', j') \neq (i, j)$ minimizes $|u_{i'j'} - u_{ij}|$ over the window.

This means that the update does not require averaging; rather it takes the form of a substitution. For this reason it is stable in the sense that the accelerated pixel value must lie in the $[w_{min}, w_{max}]$ where w_{min} and w_{max} are the minimum and maximum intensities over the window. The only computation involved is in selecting which pixel value to use in the substitution. If we take the 'window' to be the four closest pixels values $|i - i'| + |j - j'| = 1$ then this choice can be made at a cost of only two flops per pixel: At interior points in the image

Step 1 Form the x forward difference matrix $DX_{ij} = |u(i+1, j) - u(i, j)|$.

Step 2 Form the y forward difference matrix $DY_{ij} = |u(i, j+1) - u(i, j)|$.

Using DX and DY we can determine the substitution values for each pixel using three comparisons per pixel.

It has been our experience for noisy IR images that only one accelerated PGA step is needed to remove noise especially salt-and-pepper noise. The reason for this lies in the fact that for $\omega = 2$ and $n = 2$, accelerated PGA substitutes a nearby pixel value rather than averaging. This eliminates isolated noise spikes instead of reducing them by the averaging factor $1/n$ as would be the case in unaccelerated PGA with $\omega = 1$.

5 Conclusion

PDE image processing is a new science that has the potential to completely rework our understanding of images and their structural content. This paper has surveyed three methods in this area and their relations to each other. Variational approaches such as the Mumford-Shah method provide control over image approximation, smoothing and boundary extent. These methods are very successful but require significant processing time; as such they can benefit from prefiltering methods such as peer group averaging and shock filtering. The latter two methods have a number of other applications. Surprisingly they are equivalent for 1d signals even though they are derived from completely different points of view with shock filtering growing out of front propagation ideas and PGA being motivated by a desire to work with images on a discrete rather than continuous level. These three approaches should be viewed not as competitive but rather as mutually supportive with the method of choice determined by the processing needs of the particular application.

References

- [1] L. Ambrosio and V. Tortorelli, "On the approximation of functionals depending on jumps by elliptic functionals," *Comm. Pure and Appl. Math.*, 43, pp.999-1036.
- [2] L. Ambrosio and V. Tortorelli, "Approximation of functionals depending on jumps by elliptic functionals via Γ -convergence," *Boll. Un. Mat. Ital.*, 7, pp.105-123, 1992
- [3] M. J. Black and A. Rangarajan, "The outlier process: Unifying line processes and robust statistics," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, June 1994, Seattle, Washington, pp. 15-22.
- [4] A. Blake and A. Zisserman, "Visual Reconstruction," MIT Press, Cambridge, MA, 1987.
- [5] Y. Deng, C. Kenney, M. Moore, and B. Manjunath, "Peer Group Filtering and Perceptual Color Image Quantization," submitted to ISCAS 1999, Orlando FL.
- [6] Y. Deng, G. Hewer, C. Kenney, and B. Manjunath, "Peer Group Image Processing," submitted to *IEEE Trans Image Proc.* March, 1999.
- [7] D.L. Donoho and I.M. Johnstone, "Ideal De-Noising in a Basis Chosen from a Library of Orthogonal Bases," *Comptes Rendus Acad. Sci. Paris A*, 319, (1994), pp. 1317-1322.

- [8] Joel H. Ferziger, *Numerical Methods for Engineering Application*, Wiley and Sons, New York, 1981.
- [9] D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRFs: Surface reconstruction," *IEEE Trans. Pat. Anal. Mach. Intell.*, 13(5), pp. 401-412, May 1991.
- [10] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images," *IEEE Trans. Pat. Anal. Mach. Intell.*, 6, pp. 721-741, 1984.
- [11] D. Geman, S. Geman, C. Graffigne and P. Dong, "Boundary detection by constrained optimization," *IEEE Trans. Pat. Anal. Mach. Intell.*, 12, pp. 609-628, 1990.
- [12] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pat. Anal. Mach. Intell.*, 14, pp. 376-383, March 1992.
- [13] G. Hewer, C. Kenney, L. Peterson and A. Van Nevel, "PDE Techniques for Variational Image Processing," Proc. 1997 ICIP Conference, Santa Barbara, Oct. 1997.
- [14] G.A. Hewer, C. Kenney, and B.S. Manjunath, "Variational Image Segmentation Using Boundary Functions," *IEEE Trans. Image Processing*, Vol. 7, No. 9, pp. 1259-1269, 1998.
- [15] H. Li, B. S. Manjunath, and S. K. Mitra, "A contour based approach to multisensor image registration," *IEEE Trans. Image Processing*, Vol. 5, No. 4, pp. 235-245, 1995.
- [16] G. Koepfler, J. M. Morel and S. Solimini, "Segmentation by Minimizing Functionals and the Merging Methods," *IEEE Proc. of the 12th GRETSI Colloque*, 1991.
- [17] G. Koepfler, C. Lopez and L. Rudin, "Data Fusion by Segmentation: Application to Texture Discrimination," *IEEE Proc. of the 14th GRETSI Colloque*, pp. 707-710, Sept. 1993.
- [18] D. Marr and E. Hildreth, "Theory of Edge Detection," *Proc. Roy. Soc. Lond. B*, 207, pp. 187-217, 1980.
- [19] D. Marr, *Vision*, Freeman and Co., 1980
- [20] E. Abreu, M. Lightstone, S. K. Mitra and K. Arakawa, "A new Efficient Approach for the Removal of Impulse Noise from highly Corrupted Images," *IEEE Trans. on Image Processing*, vol. 5 (1996), pp. 1012-1025.
- [21] J. Morel and S. Solimini, *Variational Methods in Image Segmentation*, Birkhäuser, Boston, 1995.
- [22] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, 1985.
- [23] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," *Image Understanding*, S. Ullman and W. Richards, eds., 1988.
- [24] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," *Comm. on Pure and Appl. Math.*, Vol. XLII, No. 4, 1989.
- [25] Mary Oman, "Study of Variational Methods Applied to Noisy Step Data," unpublished, 1996.
- [26] S. Osher and L. Rudin, "Feature-oriented image enhancement using shock filters," *SIAM J. Numer. Anal.*, vol. 27, pp. 919-940, 1990.
- [27] S. Osher and L. Rudin, "Shocks and other non-linear filtering applied to image processing," *Proceedings SPIE Appl. Dig. Image Proc. XIV*, vol. 1567, pp. 414-430, 1991.
- [28] P. Perona and J. Malik, "Scale-Space and Edge Detection Using Anisotropic Diffusion," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. PAMI-12 (1990), pp. 629-639.
- [29] M. Proesmans, E. Pauwels and L. van Gool, "Coupled Geometry-driven diffusion equations for low-level vision," in *Geometry-Driven Diffusion in Computer Vision*, B.M. ter Harr Romeny, Ed., Kluwer Academic Publ., Boston, 1994, pp. 1991-228.
- [30] A. Rangarajan and R. Chellappa, "A continuation method for image estimation using the adiabatic approximation," in R. Chellappa and A. K. Jain, editors, *Markov Random Fields: Theory and Applications*, Academic Press, 1993.
- [31] Thomas J. Richardson, *Scale independent piecewise smooth segmentation of images via variational methods*, Ph.D. Thesis, Mass. Inst. Tech., Feb. 1990.
- [32] T. Richardson and S. Mitter, "Approximation, computation and distortions in the variational formulation," in *Geometry-Driven Diffusion in Computer Vision*, B.M. ter Harr Romeny, Ed., Kluwer Academic Publ., Boston, 1994, pp. 169-190.

- [33] A. Rosenfeld and M. Thurston, "Edge and curve detection for visual scene analysis," *IEEE Trans. on Computers*, C-20, pp. 562-569, 1971.
- [34] L. Rudin, "Images, Numerical Analysis of Singularities, and Shock Filters," Ph.D. thesis, Computer Science Dept., Caltech, Pasadena, CA, Tech. Report 5250:TR:87, 1987.
- [35] Scribner, et al., "Nonuniformity correction for staring IR focal plane arrays using scene-based techniques," SPIE, vol. 1308, pp224-232, *Infrared Detectors and Focal Plane Arrays* (1990).
- [36] J. Shah, "Segmentation by nonlinear diffusion," *Conference on Computer Vision and Pattern Recognition*, pp. 202-207, June 1991.
- [37] A. Witkin, "Scale-space filtering," *International Joint Conference on Artificial Intelligence*, Karlsruhe, pp. 1019-1021, 1983.
- [38] L.P. Yaroslavsky, "Linear and Rank Adaptive Filters for Picture Processing," in *Digital Image Processing and Computer Graphics: Theory and Applications*, L. Dimitrov and E. Wenger, Eds. Wien, Muenchen: R. Oldenburg, pp. 374, 1991.
- [39] L. Yin, R. Yang, M. Gabbouj and Y. Neuvo, "Weighted Median Filters: A Tutorial," *IEEE Trans. Circ. Sys. - II: Analog and Dig. Sig. Proc.*, vol. 43 (1996) pp. 157-192.
- [40] A. Yuille and T. Poggio, "Scaling theorems for zero-crossings," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8, 1986.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References RTO-EN-7 AC/323(SCI)TP/16	3. Further Reference ISBN 92-837-1021-5	4. Security Classification of Document UNCLASSIFIED/ UNLIMITED																
5. Originator Research and Technology Organization North Atlantic Treaty Organization BP 25, 7 rue Ancelle, F-92201 Neuilly-sur-Seine Cedex, France																			
6. Title Application of Mathematical Signal Processing Techniques to Mission Systems																			
7. Presented at/sponsored by the Systems Concepts and Integration Panel (SCI) and held in Köln, Germany, 1-2 November 1999, in Paris, France, 4-5 November 1999 and in Monterey, USA, 9-10 November 1999.																			
8. Author(s)/Editor(s) Multiple			9. Date November 1999																
10. Author's/Editor's Address Multiple			11. Pages 116																
12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other RTO unclassified publications is given on the back cover.																			
13. Keywords/Descriptors <table border="0"><tr><td>Signal processing</td><td>Mathematical models</td></tr><tr><td>Mission effectiveness</td><td>Integrated systems</td></tr><tr><td>Detectors</td><td>Data processing</td></tr><tr><td>Image processing</td><td>Calculus of variations</td></tr><tr><td>Wavelet transforms</td><td>Filters</td></tr><tr><td>Partial differential equations</td><td>Data compression</td></tr><tr><td>Evidence</td><td>Video signals</td></tr><tr><td>Multisensors</td><td>Communication and radio systems</td></tr></table>				Signal processing	Mathematical models	Mission effectiveness	Integrated systems	Detectors	Data processing	Image processing	Calculus of variations	Wavelet transforms	Filters	Partial differential equations	Data compression	Evidence	Video signals	Multisensors	Communication and radio systems
Signal processing	Mathematical models																		
Mission effectiveness	Integrated systems																		
Detectors	Data processing																		
Image processing	Calculus of variations																		
Wavelet transforms	Filters																		
Partial differential equations	Data compression																		
Evidence	Video signals																		
Multisensors	Communication and radio systems																		
14. Abstract <p>Presents a whole range of perspectives for different levels of mathematical signal processing, based on some of the most promising techniques. Particular attention is paid to the following subjects:</p> <ul style="list-style-type: none">• Wavelet analysis: summary of the possibilities; application to detection in natural background radiation and extraction of primitive invariants.• The concept of Multirate Filter Banks in conjunction with the various transforms which this technique enables; applications to compressed video image and sequence transmission, to noise rejection, to jamming and to encoding.• Variational methods based on partial derivative equations for image processing and multi-scale video sequences; presentation of different image segmentation approaches;• Multi-sensor processing based on the theory of evidence: processing of the functions of detection, classification, matching of ambiguous observations, or tracking, with the aim of solving problems such as data modelling, decision making, the management of non-uniform reference systems, or the integration of contextual knowledge.																			



RESEARCH AND TECHNOLOGY ORGANIZATION

BP 25 • 7 RUE ANCELLE

F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE

Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DIFFUSION DES PUBLICATIONS

RTO NON CLASSIFIEES

L'Organisation pour la recherche et la technologie de l'OTAN (RTO), détient un stock limité de certaines de ses publications récentes, ainsi que de celles de l'ancien AGARD (Groupe consultatif pour la recherche et les réalisations aérospatiales de l'OTAN). Celles-ci pourront éventuellement être obtenues sous forme de copie papier. Pour de plus amples renseignements concernant l'achat de ces ouvrages, adressez-vous par lettre ou par télécopie à l'adresse indiquée ci-dessus. Veuillez ne pas téléphoner.

Des exemplaires supplémentaires peuvent parfois être obtenus auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus sur la liste d'envoi de l'un de ces centres.

Les publications de la RTO et de l'AGARD sont en vente auprès des agences de vente indiquées ci-dessous, sous forme de photocopie ou de microfiche. Certains originaux peuvent également être obtenus auprès de CASI.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE

Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

BELGIQUE

Coordinateur RTO - VSL/RTO
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evère, B-1140 Bruxelles

CANADA

Directeur - Recherche et développement -
Communications et gestion de
l'information - DRDCGI 3
Ministère de la Défense nationale
Ottawa, Ontario K1A 0K2

DANEMARK

Danish Defence Research Establishment
Ryvangs Allé 1, P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE

INTA (RTO/AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

ETATS-UNIS

NASA Center for AeroSpace
Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320

FRANCE

O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GRECE

Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

HONGRIE

Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
H-1525 Budapest P O Box 26

ISLANDE

Director of Aviation
c/o Flugrad
Reykjavik

ITALIE

Centro documentazione
tecnico-scientifica della Difesa
Via Marsala 104
00185 Roma

LUXEMBOURG

Voir Belgique

NORVEGE

Norwegian Defence Research
Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

PAYS-BAS

NDRCC
DGM/DWOO
P.O. Box 20701
2500 ES Den Haag

POLOGNE

Chief of International Cooperation
Division
Research & Development Department
218 Niepodleglosci Av.
00-911 Warsaw

PORTUGAL

Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

REPUBLIQUE TCHEQUE

VTÚL a PVO Praha /
Air Force Research Institute Prague
Národní informační středisko
obránného výzkumu (NISCR)
Mladoboleslavská ul., 197 06 Praha 9

ROYAUME-UNI

Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

TURQUIE

Millî Savunma Başkanlığı (MSB)
ARGE Dairesi Başkanlığı (MSB)
06650 Bakanlıklar - Ankara

AGENCES DE VENTE

NASA Center for AeroSpace
Information (CASI)

Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320
Etats-Unis

The British Library Document
Supply Centre

Boston Spa, Wetherby
West Yorkshire LS23 7BQ
Royaume-Uni

Canada Institute for Scientific and
Technical Information (CISTI)

National Research Council
Document Delivery
Montreal Road, Building M-55
Ottawa K1A 0S2
Canada

Les demandes de documents RTO ou AGARD doivent comporter la dénomination "RTO" ou "AGARD" selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et AGARD figurent dans les journaux suivants:

Scientific and Technical Aerospace Reports (STAR)

STAR peut être consulté en ligne au localisateur de
ressources uniformes (URL) suivant:
<http://www.sti.nasa.gov/Pubs/star/Star.html>
STAR est édité par CASI dans le cadre du programme
NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
Etats-Unis

Government Reports Announcements & Index (GRA&I)

publié par le National Technical Information Service
Springfield
Virginia 2216
Etats-Unis
(accessible également en mode interactif dans la base de
données bibliographiques en ligne du NTIS, et sur CD-ROM)



Imprimé par le Groupe Communication Canada Inc.
(membre de la Corporation St-Joseph)
45, boul. Sacré-Cœur, Hull (Québec), Canada K1A 0S7



RESEARCH AND TECHNOLOGY ORGANIZATION

BP 25 • 7 RUE ANCELLE

F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE

Telefax 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DISTRIBUTION OF UNCLASSIFIED
RTO PUBLICATIONS

NATO's Research and Technology Organization (RTO) holds limited quantities of some of its recent publications and those of the former AGARD (Advisory Group for Aerospace Research & Development of NATO), and these may be available for purchase in hard copy form. For more information, write or send a telefax to the address given above. **Please do not telephone.**

Further copies are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO publications, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your organisation) in their distribution.

RTO and AGARD publications may be purchased from the Sales Agencies listed below, in photocopy or microfiche form. Original copies of some publications may be available from CASI.

NATIONAL DISTRIBUTION CENTRES

BELGIUM

Coordinateur RTO - VSL/RTO
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evêre, B-1140 Bruxelles

CANADA

Director Research & Development
Communications & Information
Management - DRDCIM 3
Dept of National Defence
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC

VTÚL a PVO Praha /
Air Force Research Institute Prague
Národní informační středisko
obraného výzkumu (NISCR)
Mladoboleslavská ul., 197 06 Praha 9

DENMARK

Danish Defence Research
Establishment
Ryvangs Allé 1, P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE

O.N.E.R.A. (ISP)
29 Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GERMANY

Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

GREECE

Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

HUNGARY

Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
H-1525 Budapest P O Box 26

ICELAND

Director of Aviation
c/o Flugrad
Reykjavik

ITALY

Centro documentazione
tecnico-scientifica della Difesa
Via Marsala 104
00185 Roma

LUXEMBOURG

See Belgium

NETHERLANDS

NDRCC
DGM/DWOO
P.O. Box 20701
2500 ES Den Haag

NORWAY

Norwegian Defence Research
Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

POLAND

Chief of International Cooperation
Division
Research & Development
Department
218 Niepodleglosci Av.
00-911 Warsaw

PORTUGAL

Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

SPAIN

INTA (RTO/AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

TURKEY

Millî Savunma Başkanlığı (MSB)
ARGE Dairesi Başkanlığı (MSB)
06650 Bakanlıklar - Ankara

UNITED KINGDOM

Defence Research Information
Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES

NASA Center for AeroSpace
Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320

SALES AGENCIES

**NASA Center for AeroSpace
Information (CASI)**

Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320
United States

**The British Library Document
Supply Centre**

Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom

**Canada Institute for Scientific and
Technical Information (CISTI)**

National Research Council
Document Delivery
Montreal Road, Building M-55
Ottawa K1A 0S2
Canada

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)

STAR is available on-line at the following uniform resource locator:

<http://www.sti.nasa.gov/Pubs/star/Star.html>

STAR is published by CASI for the NASA Scientific and Technical Information (STI) Program

STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
United States

Government Reports Announcements & Index (GRA&I)

published by the National Technical Information Service

Springfield

Virginia 22161

United States

(also available online in the NTIS Bibliographic

Database or on CD-ROM)



Printed by Canada Communication Group Inc.

(A St. Joseph Corporation Company)

45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7